The potential of ontologies for the empirical assessment of machine learning techniques in operational oceanography

Enrique Wulff [1]

[1] *Marine Sciences Institute of Andalusia, Spanish National Research Council (CSIC) Campus del Río San Pedro Cadiz 11510, Spain*

Abstract

A role for ontologies is key for the digital transformation of operational oceanography processes to the adoption of artificial intelligence and machine learning. Marine ontologies, a common concept among these tools, can lead to lower costs and more flexibility in identifying and classifying marine data. This study explores a demonstration that proves the potential of ontologies to fulfill the requirements outlined in the case of how to visualize computer datasets. A selective network of records, including visual and textual features that can be annotated from video and image sequences, with subsea parameters as the target of interest. The sample is divided into ontology and machine learning (ML) datasets to predict the importance of data visualization methods. The predicted suitability is strong with data classification that belongs to the machine learning dataset. However, the initial results from the study are encouraging, because ontologies' tools are proposed as automatic reasoning mechanisms. This proof of principle shows that it is almost guaranteed that marine ontologies can be built to make visual patterns for marine data usable by different communities, which could be used to identify "interesting" functions at the intersection of computer vision and machine learning in general.

Keywords

Ontology, machine learning, artificial intelligence, data visualization, classification

1. Introduction

An improved ontological representation of marine data as a paradigm for pattern analysis software development requires more work on combining different modes of inference (OWL, ML), the design of algorithms for data classification (DC) and visual data recognition (DR) for signal and image analysis [1]. This poses the problem of how should marine databases be represented. An ontology of a domain is an "explicit formal specification of the terms in the domain and relations among them" [2]. An ontology fully describes the subject area as a dictionary, in a way it is the ideal tool when we focus on the generation of contextual descriptions for images (in 3D shape retrieval for example [3]). Most of pattern analysis algorithms in oceanography, are to be used for object detection and recognition research, motivated by this challenge it can be proved that an ontology could be a relevant approach to the problem of marine data recognition and classification.

The marine data received from wireless sensor networks are heterogeneous in nature. For instance, the existing marine acoustic data cannot meet the amount of data required for training models [4]. In particular, positioning and orientation systems, and other sensor technology, is based on multi-beam echo sounder system acceptance and quality assurance. An automated system producing multiple overlapping range images that was the first for correctly registered mapping of the ocean floor [5]. Whether data come from GIS technology, the Web or any other present or future approach they share common ground [6]. A role for ontologies is key in the development of application software for the acquisition, analysis and display of real time marine data, for the generation of model scenario databases for their retrieval, and display at the time of an event and for the decision support systems following a

Proceedings of the International Conference on Biomedical Ontologies 2023, August 28th-September 1st, 2023, Brasilia, Brazil
EMAIL: enrique.wulff@icman.csic.es
ORCID: 0000-0001-8104-6147

standard procedure [7-8]. To define and develop intelligent systems, has been proposed in recent times, giving a rise in both precision and recall as well as facilitating system interoperability through data harmonization [9-12]. Ensuring interoperability between marine databases is a huge challenge. Terms and codes used to structure exploit the data comes from many sources and are continuously evolving.

The problem pattern analysis (PA) is facing consists in finding an adequate visualization, a "good" figure, since humans are only capable of perceiving objects in at most three dimensions [13]. This means that pattern analysis has to find a method to reduce the heterogeneity of the set of data under study, thus allowing an analysis of the problem of stability of pattern. For practical reasons usually only recognition and classification of those data are allowed (best practices must be carried out by focusing on structure and naming consistency) [14]. Image recognition tasks are at the centre of the ongoing machine learning revolution, an approach that in the monitoring of coastal seas is focused on using automated classification algorithms based on random forest or deep learning approaches [15]. However, the field of marine image processing lacks the large numbers of annotations in images required [16]. The lack of correspondence between the visual representation of the image and its meaning calls for the performance of Machine Learning, expressed through semantic resources such as ontologies.

The problem of trying to solve the visual parameters of images or videos focuses on tasks such as object detection, data recognition, and multi-level data classification. Such an example could be that of studying how the air and sea interact with each other during El Niño/La Niña onsets, by using pattern analysis with ocean data assimilation techniques [17]. This is an issue where content-based image retrieval is approached in terms of Machine Intelligence [18][19]. As such Pattern Analysis and Machine Intelligence (PAMI) is an element of scholarship proposed in the last thirty years and where it has been a continuous need to develop new data recognition and classification methods and advanced equipment for solving modern practical problems [20].

1.1. Pattern Analysis [and Machine Intelligence (PAMI)] and Marine Ontologies

Rethinking pattern analysis of marine data means to investigate the rich variety of application scenarios offered by marine ontologies. While otherwise adding value to public data using semantic web axioms and machine learning to support annotation contribute to pose and solve issues involving ocean data classification.

Application of ontologies in ocean data grows out of an Artificial Intelligence (AI) engagement with marine data metrics of interoperability and reuse. Ontologies serve as such a tool and method to assess the added value robotic technology brings into the marine environment (autonomous underwater vehicles (AUVs) or (ocean floor observation systems) OFOSs). From a pattern recognition point of view, ontologies for describing sensors and sensor networks work in the context of Sensor Web applications. Knowledge representation in the Internet of Things (IoT) presents a general architecture of Sensor Web applications. And that is why it provides huge numbers of interconnected data across an extended variety of various ocean regions, which classifications depend on the specific context and resources of LinkedData.

By using ontological representation, the best of technical progress, undertaken by a community to unambiguously set definitions and interconnect concepts in various field, is captured. The use of ontologies for representing database entities has proven to be advantageous in the field of Pattern Analysis and Machine Intelligence (PAMI) (see Table 1).

Table 1
The main features provided by ontologies in support of PAMI

Ontology feature	Utility in PAMI
Classes and relations	When ontology reasoning is applied to sensor data, rdf:type will be connected to a class name of an ontology
Domain vocabulary	Ontologies provide a domain vocabulary that can be exploited to create a dense network of relationships among the entities, and serve software applications, and GIS
Metadata and descriptions	Biodiversity data, especially in marine domain, have database entities represented as ontologies where these last are primarily used for metadata that describe raw data providing contextual information
Axioms and formal declarations	Ontology axioms and applied reasoning on them are related to the recognition of object presence in a time interval

The concept of marine ontologies may be the solution in developing systems and workflows that would meet the various possible marine data requirements and from them derive up to standard products/maps without human assistance except at the user interface. As shown in Figure 1, research on ontology topics can be followed from different perspectives. The index is the percentage of the publications in the ontology sub-areas of research. It covers semantic web, web services and so forth. Especially, the semantic web, data integration, and web service have attracted the attention of a large number of researchers in recent years while the research on the topics of data source, relation extraction and heterogeneous data seems less consistent. One element is the major cause of these problems, as far as a common ontology for marine data is necessary to enable exchange and integration of data. Terminology is used to describe similar data can vary between marine specialties or world ocean regions, which can complicate data searches and data integration.

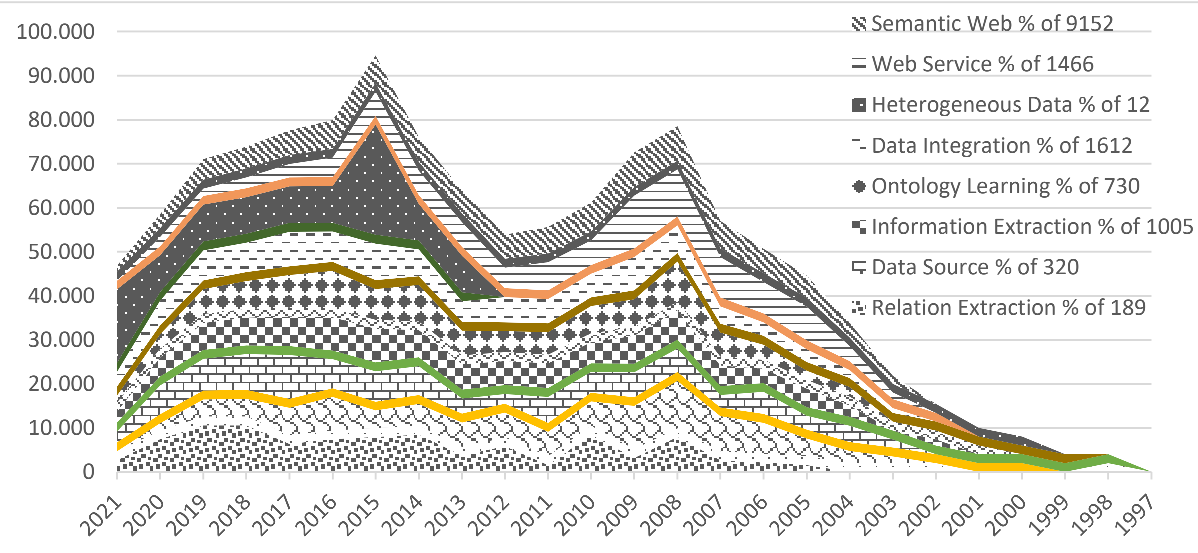

Figure 1: Ontology subareas of research (dependent variable: percent of publication in ontology sub-areas)

The ontology-based research illustrates especially how those involved with marine data should be informed about marine ontology developments. Opportunities to enhance their development will contribute to the success of ontology in the way that certain concepts and ideas start to unfold. It is customary to consider emerging observational patterns making sense out of methods that captures concepts leading to finding out what is visible. To conduct this insight, for example in coastal web atlases (CWA), developers should intensify future efforts to improve data discovery, sharing, and integration on the base of ontologies.

1.2. Ontologies and Marine Robotics

Marine data classification has been studied widely in the field of marine robotics; while pattern recognition is a process of finding regularities and similarities in data using machine learning data which is the perspective of marine robotics. Marine robotics has undergone a phase of dramatic increase and its quantitative landscape, status quo and current workflow is shaped by its own pattern analysis, data recognition and classification issues.

From the point of view of marine robotics key issues in ocean data management concern two different PAMI realities representing, detecting, and tracking features and the process of integrating real sensor data with a model of an ocean process. As a standard knowledge representation ontology can facilitate the development of these marine robotic applications in various ways:

- Providing a consistent set of terminology (domain vocabulary), and concepts in the robot's knowledge representation (definitions, relations, domain axioms and taxonomy)
- Enabling design pattern guidelines for content analysis of complex tasks, environment, etc.
- Ensuring a common repository of knowledge that can be shared among various robotic systems
- Highlighting more efficient new relations through the analysis of data generated using ontologies

1.3. Contribution of this paper

The purpose of this paper is to identify relevant pattern analysis research in marine data classification and recognition, and to review its intersection with the state-of-the art in marine ontologies. It focuses on the 3D modeling and analysis domain, computer vision and interactions are described for machine learning (ML) and marine ontologies.

2. Method

All the R&D efforts in pattern analysis, classification and recognition of data have been kept rising over the current period (1991 to 2021). To obtain a general understanding of this research question concerning marine data we systematically reviewed the IEEE Pattern Analysis and Machine Intelligence, IEEE Access, IEEE Journal of Oceanic Engineering, Sensors, and Information Visualization. Initially, we identified the appropriate subset of articles from these conferences and journals. We then conducted an in-depth qualitative analysis of the relevant work, re-removing and refining the characteristics of the marine data interaction of PA. The histogram theory inspired us to take a general approach to this analysis, which systematically analyzes the data until significant categories appear. This methodological approach is based on define and refine categories based on a representative set of qualitative data, here are documents that are then used to progressively build a theoretical model. This approach has been used in pattern analysis and related areas such as data classification and data recognition before, and recognized its role for the importance of establishing a much-needed theoretical framework for visualization.

2.1. State-of-the-art of Pattern Analysis and data classification and recognition in marine data

We started our efforts with a carefully selected list of important publications in interactive machine learning and marine data. Using these candidate documents, we first tried an open approach to coding to identify "interesting" features at the intersection of computer vision and machine learning in general. Although this resulted in a high-level structure [19], it was impractical to make the analysis more concrete. Therefore, we decided to analyze a much larger set of sample articles, with two implications for our methodological options. (1) We understood the need to look at specific pattern analysis problem (in our scenario, intelligent driving, image synthesis, and object pose measurement) searching to make the research more focused, practical, and concrete. (2) We needed automated methods to narrow down the pool of potentially interesting articles.

During this process, we repaired the retrieval practice in cleaning its criteria and coding selection multiple times. Our final workflow consisted of four main steps, shown in Figure 2: 1.) Obtaining ontology research trends, 2.) Reviewing the previous research and application of pattern analysis, 3.) Identifying marine data classification and recognition issues, 4.) Searching for pattern analysis and machine learning parameters to encode for a large part of the ontologies' semantic content.

2.2. Sample network of records

Our common goal was the ontology research developed and how its implementation interacts in the pattern analysis and marine data communities. We decided to take a representative sample of papers, made up of every paper ever published in a Web of Science (WoS) source titles in the marine pattern analysis community from 1991 to 2021. From the database (WoS), is defined a collection of pattern analysis (PA) and marine data records (6048), data classification (DC) and marine data papers (3242) and data recognition (DR) and marine data records (1214), for a total of 9,899 records.

2.2.1. Paper metadata-based filtering

Methodological options were driven by the idea that the state of the art of ontology (machine/deep learning) research could be determined by using metadata. By metadata, we refer to aspects of the words-in-title that were deemed essential to facilitate a meaningful analysis in a full-content context. The initial synthesis was accomplished by deciding on a uniform list of metadata and their distribution along the years, as found in Figure 1. Based on this metadata definition, we implemented metadata lists from the sets of records in PA and data classification and recognition. The final metadata lists and statistics from this metadata filtering process are provided in Figures 2a,b,c.

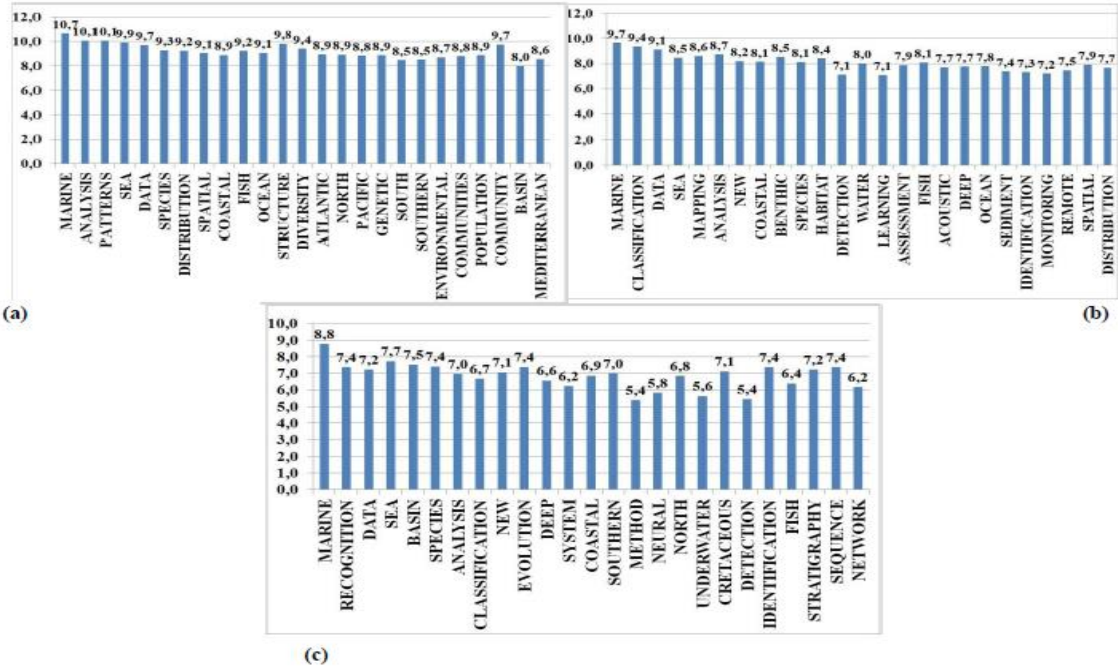

Figure 2: Comparison of metadata efforts in tracing WoS records to measure pattern analysis, marine data classification and marine data recognition shown in three log-scale histograms; 25 top metadata required with pattern analysis capture the data to detect, recognize and identify target of interests from physical, optical, fluid, and chemical underwater parameters (a); histogram estimated by image or video parameters with an emphasis on multilevel data classification is reported by its 25 top metadata (b); ranking title terms (metadata) of the documents on data recognition (c)

We formed a set of primary papers in marine ontologies with and without the initial criteria on PA and machine intelligence for data classification and recognition. Not all of these metadata allow to express the semantic content of an image. The discrepancy between the visual presentation of the image and its meaning requires machine learning performance expressed in terms of semantic resources as a ontology. So from the data set, including records in PA and corresponding DC and DR values (9899) are extracted two test sets on ontologies (42) and machine learning (210). In this way, the use of marine ontologies as the data classification and recognition technique focuses on the viability of using ontologies to solve the problem of pattern analysis.

2.2.2. Manual and automatic sample check

The 42 ontological papers were manually checked using the following criteria. First, we checked if the paper is a theoretical and evaluative framework or if it deals with a combination of applied or technical visual methods; as we planned to build a theoretical model for visualization. Second, we checked whether the paper addresses the combination of pattern analysis (PA), data classification (DC) and data recognition (DR), and whether the interaction returns to the visualization area. This had the advantage to present an interesting one workflow for the multi-source, multi-format, multi-dimension characteristic of marine data. Moreover, there is a return to the visualization area in its framework design that considers underlying data patterns. Given our focus on visualization we include this model that even feedback to the analysis of the 3D marine data. One major advantage of this method is its ability to define a semantic model of the issue under scrutinize (PA, DC & DR) combined with the associated domain of visualization to list the data visualization theories brought by marine data and observations, that range from the digital transformation of operational oceanography processes to the adoption of artificial intelligence. On this basis, we manually analyzed the first 42 candidate relevant documents obtained. Table 2 provides a partial list of the 42 specific ontological contexts detected in the PA and DC and DR data sources, and the extent to which they provide the ontology tools they use.

Table 2
List of the ontological contexts detected in the PA and DC and DR data sources (A/B: applied//theoretical; DV: data visualization feedback)

Num	Experimental Organism	A/B	DV	Ontology
1	Dynamena pumila	B	Y	Gene Ontology (GO) and KEGG pathway
2	Takifugu rubripes	A	Y	Gene Ontology (GO)
3	phytoplankton	A	Y	Gene Ontology (GO)
4	nd	A	Y	nd
5	Dreissena polymorpha	B	Y	Gene Ontology (GO)
6	Atlantic salmon	A	Y	Gene Ontology (GO) and UniProt Knowledgebase
7	Micromonas polaris; Pyramimonas tychotreta	B	Y	Gene Ontology (GO)
8	Crassostrea gigas	A	Y	Gene Ontology (GO)
9	Nd	B	Y	Genomic Standards Consortium's MIxS and Environment Ontology (ENVO); EMP Ontology (EMPO) of microbial environments
10	Chlamys farreri	A	Y	Gene Ontology (GO) and Eukaryotic Orthologous Groups (KOG) and Kyoto Encyclopedia of Genes and Genomes (KEGG)
11	Nd	B	Y	Protégé environment (ontology)
12	Nd	B	Y	Protégé environment (ontology)
13	Eucheuma denticulatum	A	Y	Gene ontology (GO)
14	48 species of freshwater and marine fish	A	Y	Gene ontology (GO)
15	Larimichthys crocea	A	Y	Gene ontology (GO)
16	Seriola lalandi	B	Y	Gene ontology (GO)
17	marine and FW sticklebacks	B	Y	Gene ontology (GO)
18	Genypterus chilensis	A	Y	Gene ontology (GO)
19	Ceriops	A	Y	Gene ontology (GO)
20	Human	A	Y	Gene ontology (GO)
21	Zostera muelleri	B	N	Gene ontology (GO)
22	Nd	B	Y	Integrated Ocean Observatory System ontology and the Marine Metadata Interoperability ontology (MMI)
23	Nd	A	Y	SWEET (Semantic Web for Earth and Environmental Terminology)
24	Mytillus galloprovincialis; Crassostrea-gigas; Chlamys farreri	A	Y	Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG)
25	Human	B	Y	BRENDA Tissue Ontology (BTO); Gene ontology (GO); human anatomy atlas CAVEman
26	Tachypleus tridentatus	A	Y	Gene ontology (GO)
27	Nd	B	Y	Uberon metazoananatomy ontology; Environmental Ontology (EnvO)
28	Orcinus orca	B	N	Gene ontology (GO)
29	Perna viridis; Mytilus galloprovincialis; Manila clam	B	Y	Gene ontology (GO)
30	Nd	B	N	nd
31	Pinctada martensii	A	N	Gene ontology (GO)
32	Mytilus galloprovincialis	A	N	Gene ontology (GO)
33	Sea cucumbers	B	Y	Gene ontology (GO)
34	Epinephelus coioides	A	Y	Gene ontology (GO)
35	Sparus aurata	B	Y	Gene ontology (GO)
36	Drosophila	B	Y	Gene ontology (GO); Kyoto Encyclopedia of Genes and Genomes (KEGG)
37	Mytius galloprovincialis	A	Y	Gene ontology (GO)
38	mouse brain	A	Y	Gene ontology (GO); KEGG pathway; web tool DAVID
39	Pacific oyster Crassostrea gigas	A	Y	GENE ontology (GO); program KAAS (KEGG Automatic Annotation Server)
40	adult male mice	A	Y	Gene ontology (GO); web tool DAVID (The Database for Annotation, Visualization and Integrated Discovery
41	dogfish shark (Squalus. acanthias)	A	Y	Gene ontology (GO); PANTHER gene ontology classification system (Applied Biosystems)
42	Nd	B	N	nd

After an automated process, the set of 210 papers corresponding to machine learning was filtered based on the fact that one of the most frequently used data visualization techniques in machine learning is the histogram plot. ML-based data visualization techniques were approached through metadata generated from a base histogram and classified into four levels: disseminative, observational, analytical and model-developmental. That is to say, a theoretical framework, because a visualization technique that builds on machine learning therefore attests its power for interactive analysis of heterogeneous marine data, it can deliver relevant pattern analysis content in the appropriate mode. Table 3 lists these visualization levels. Through boxplot with the ontological and machine learning (ML) datasets, we found differential expression of how their values are spread out and detect schematically their outliers (Figure 3). The temporal analysis parameters for machine learning (ML) are listed in Table 4.

Table 3
Levels of data visualization in machine learning methods for pattern analysis (marine data) (metadata generated from a base histogram

	analytics	association	case study	classification	comparison	complex	correlation	development	information	mapping	model	monitor	observation	pattern	prediction	recognition	regression	sensor	simulation	time series	other	SUM
Papers	1	1	4	70	7	3	1	4	2	16	20	12	2	13	13	7	1	12	3	2	16	210
disseminative				1					2			4		1						2	3	13
observational			1	11								2	2	1				12			4	33
analytical	1	1	3	34	7	3	1					6		10			1				3	70
model-developmental				24				4		16	20			1	13	7			3		4	92

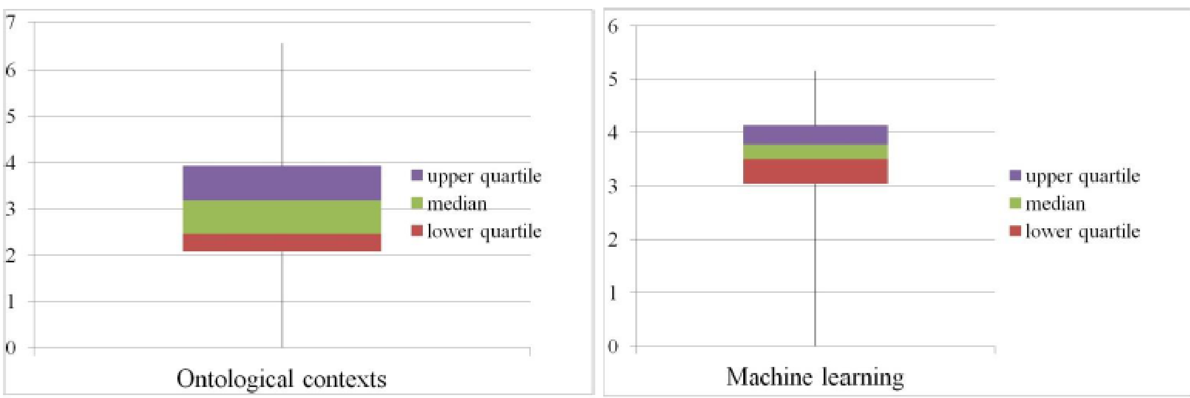

Figure 3: Differential expressions of relevance from the Ontological and Machine learning (ML) datasets (dependent variable: ln(cit))

Table 4
Machine learning for pattern analysis time statistics (marine data)

	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	SUM
Papers	3	3	7	5	4	6	9	12	16	15	27	55	48	210
disseminative							1		1	1	1	5	4	13
observational	1		1			1		1			6	12	11	33
Analytical	1	1	3	2	1	4	4	6	6	6	7	14	17	70
model-developmental	1	2	3	3	3	1	4	5	9	8	13	24	16	92

3. Results and discussion

3.1. Ontology research trends review

The outputs of review using the ontology research trends are shown in Figure 1. As mentioned in Section 2.1, database entities represented as ontology terms result in a rich variety of scenarios that store in annotations their features and strengths. The ultimate goal is a system that combines visual and textual semantics to regularly annotate video sequences final aim is a system that will link the visual and text semantics in order to routinely annotate video sequences with the appropriate keywords of a domain expert. Most ontology-based cognitive vision promising results occurred in 2008.

By 2015, heterogeneous marine data, believed by visualization techniques to be of strategic importance, had their top priority. It is also the subject of scientific research, as evidenced by a large number of research papers, books, and reports. Highlights were the initiative to create in marine systems well-founded ontologies embedding these domain semantic and logical frameworks in the underwater environments thus providing opportunities for intelligent observatory units. The details of communication ontology that can be used by Remotely Operated Vehicles (ROVs) to transmit data and commands between vehicles and operators is defined by OWL in SWARMs platform. SWARMs users can estimate the rate of spread of pollutants and determine the level of pollution and the estimated size of the polluted area. In marine biology, search engines do use ontologies like SWEET for engaging the coral reef research community via a cyberinfrastructure network.

3.2. State of the art of pattern analysis in marine data classification and recognition

The outputs selection was initially based on the idea that machine learning (ML) enhanced by the ontology is able to compare pattern analysis performances using marine data. But ontologies specific coverage statistics are few, and it is difficult to say what actually constitutes a significant part of the terms in an ontology.

Typically, a generic ontology design pattern is developed for data from observations on the Semantic Web by unlocking the potential of compositional definitions, proposed to distinguish reliable relations for pattern analysis expression. These definitions are the necessary information to start with, because they are partitioned into mutually exclusive cross-products sets, many of which reference other candidate ontologies for chemical entities, proteins, biological qualities and zoological entities. An example of such a case is the Environment Ontology (ENVO) which, using the expressivity of OWL, grows in acceptance and participation in new user communities, thus offering an example of ontology's classes increased granularity in their logical definitions, allowing more flexibility in semantically advanced questions, inferences and analysis.

Ontologies such as the Extensive Observation Ontology (OBOE), Observations and Measurements (O&M), Semantic Sensor Network (SSN), and SWEET can be interconnected and expanded to include additional concepts more specific to the field of remote sensing, including the basic concepts that remote sensing professionals rely on to interpret remote sensing images (e.g., concepts, associated with spectral bands, spectra or texture indices). Examples of such a remote sensing ontology have already been applied, but have not yet been used in any upper ontology.

In this sense, under a definition of data classification as a process of clustering these data into a series of groups or categories, regardless of the method used for this purpose, research on pattern analysis finds a framework for marine data based on ontologies as an active output for computer vision. From this framework for an ontology-based ocean image classification that describes how to create ontological models for low- and high-level features, classifiers and rule-based expert systems, a much larger set of sources appeared.

3.3. Histograms to bridge the semantic gap between notions of content and similarity

The results from the previous section are used to build a sample of filtered dataset (9,899 records). Firstly, such sample contains metadata from thousands of word-in-title terms that can capture both concrete and abstract relationships between salient visual properties. Subsequently, histogram analysis methods were employed to compare the semantic effort by considering the metadata weight as generated on the base of global citation scores. This result in three referential frames, is shown in Figure 2a,b,c.

Structure (292 records) and communities (199) have a significant score in pattern analysis, indicating that these two parameters cause a positive effect on modelling required to discriminate relevant from non-relevant images (Figure 2a). Basin has a weak score due to the indirect value it has with the general purpose retrieval of features constructs such as predicates, relations, conjunctions, and a specification syntax for image content (for instance, photographic images).

The following is the histogram for data classification (see Figure 2b), which shows that mapping (221 records) is the main technique to classify marine data, while spatial (93) data are by now less rigidly circumscribed. The complete histogram of the attributes obtained for quantify the features in the data recognition domain is also shown Figure 2c. The importance of data recognition for evolutionary biologists (53 records) is enclosed within the scope of the study of species (67 records). The close relation with classification (61 records) ensures that a visual language can use an important mechanism for conscious control, limiting the range of possible configurations of functions that must be taken into account when performing a visual recognition task.

3.3.1. A framework for interactive visual analysis of heterogeneous marine data

The statistical quantized histogram metadata analysis was based on the available PA, DC & DR data, and focused on expressing the multi-dimension spatiotemporal marine data in one workflow. Based on the data we processed, two visualization methods are explored: ontology and machine learning (ML). The basic idea is shown in Fig. 3.

In this way, according to the data value and methodological choice, two different data classification and recognition scenarios in pattern analysis can be compared. Under the first scenario ontology-based for image retrieval and annotation was used to derive marine data patterns. Owing to the substantial positive bias in ontological feedback to the domain of visualization (Y=85.7%; Table 2), subsequent approaches for visual events were larger than in the case of the second scenario (machine learning (ML)), but because citations were restricted to 70% of the available ML data, the resulting lower quartile of 0.45 reached a best score for ML than for ontologies (0.38).

Therefore, machine learning (ML) decisions based on marine data assessment outperformed ontology-driven coding for image classification. And that, in spite of ontology mapping for underwater IoT (IoUT) supports better interoperability protocols in the context of computer vision.

3.3.2. PA, DC & DR identified through ontologies for marine data

The Table 2 is built based on the 42 selected ontology records. This is a general-purpose scheme designed for filtering the typology of the data sources. As described by the PubMed database there is a strong proximity between applied (or technical) and theory (evaluative, comparative, lessons) contents when they are both expressed in percentage terms (47.61% vs. 52.35). This analysis is then extended utilizing the feedback of each source to the domain of data visualization. It is overly positive (85%), as estimated by subject headings including in MeSH for each source.

For all the sources that have been used in this study, the ontology tools are listed in Table 2. Researchers proposed new models to cope with marine transcriptome/genome identification (80%; Table 2). They assumed Gene ontology (GO) approaches to model knowledge on the experimentation organisms. Some approaches focus on a data visualization package (Illumina sequencing technology) to provide refined descriptions of the whole scenario (1,5,18,19,24,36; Table 2). In one source (39; Table 2), the authors propose an automatic reasoning mechanism to deal with uncertainty in a quasi-empirical model using KAAS automatic annotation server. A number of marine or environmental ontologies (MMI, ENVO, EMPO, SWEET) are found (9,11,12; Table 2), they are used as dictionary learning in microbial environments to find out unavailable variables. Other ontologies' tools are also used (BRENDA Tissue Ontology (BTO)(25; Table 2), Protein Annotation through Evolutionary Relationships (PANTHER) (17,41; Table 2).

3.3.3. Machine learning levels of visualization and their temporal perspective

In this section, it is investigated whether a data visualization level can give a prediction of its suitability for a particular machine learning task. There exists a spectrum of different steps of visualization ranging from high abstraction levels (e.g., model-developmental tools) to lower levels (e.g., operational aids) (see Table 3; 210 records). To enhance this theoretical framework performance, the ML-based data visualization techniques are used based on 20 metadata, assuming that the marine data papers are categorized from different histograms which are quite reliable.

On Table 3, the level of visualization importance for suitability prediction is shown. As the first basic task of knowledge discovery, it proposes the use of data classification tools (33%); most visual analytics processes reported in PA, DC & DR literature operate at this level. The analyst knows or assumes the model to be correct only in 6% of the sources. Only in 16% human analysts need to use visualization to observe data routinely. Human analysts are able to observe input data in conjunction with the machine's "understanding" in many ways (33%). A line-up of model developmental tools gains a new understanding in terms of complexity (44%).

We can further derive from Table 3 that deep learning is the main automated support of marine data analytics using machine learning (ML) techniques, perhaps this is caused by the success of deep learning in computer vision tasks (eg. image classification, object detection, instance segmentation). In data visualization most of the deep learning studies focus on model developmental aids (44%), followed by observational tools (30%), investigative (18%) and presentational (8%) aids. Analytical and model-developmental visualization levels were shared equally among other ML techniques employed (transfer learning, ensemble methods, clustering).

As shown in Table 4, a sequence has been used to encode the data sources as the number of papers published by year. Years 2020 (55) and 2021 (48) are peaks. We can see clearly that 50% of the papers were all published in the last two years. This is not strange since the idea that in machine learning (ML) pattern analysis is gaining future, is expressed again by the importance of the two last years for the four levels of visualization (relative importance of 69%, 70%, 44%, 43%). As expected from the results in Table 3, there are gaps in both disseminative and observational data sets (with 6 years breaks in between).

4. Conclusions

This proof of principle study explores the potential uses of ontologies to encode for marine data pattern analysis literature. This study focuses on characterizing marine ontologies to select data visualization techniques. The underlying assumption is that the application of ontologies in marine data poses the problem of how should marine databases be represented. Therefore, the validation against pattern analysis in oceanography should be first put in terms of data interoperability. Using this approach could provide experts with a tool and method where they can rate ocean technologies and how they have been received in the communities where they have been placed. A data histogram approach has been adopted, which draws on the analysis of literature until significant categories appear. This has demonstrated its worth in pattern analysis, data classification and data recognition, and is regarded as an ingredient of the new generation of theoretical frameworks for data visualization. The results of the model to predict what the encoding for a large part of the ontologies' semantic content is going to look like in the future show that marine ontologies specific coverage statistics are few. It is acknowledged that the biomedical Gene Ontology (GO) currently represents the most successful implementation of ontologies in the domain of oceanography for pattern analysis applications including data visualization. It is recognized that, for machine learning data visualization, marine data scoring solutions were better than ontology-based coding for image classification. This approach has led to accurate predictions of the level of visualization importance for the example of data classification. Over the machine learning techniques most used for computer vision tasks with marine data, the result of the study outstands for it is clearly stated that deep learning is a promissory approach to gain new understandings in terms of data visualization tools. The results of this study show the potential use of marine data for pattern analysis assessment and prediction of the level of data visualization. This method shows the potential of ontologies to support the generation of model scenarios for image retrieval and annotation, and to aid for the empirical assessment of machine learning techniques. A single example data visualization was used as an application for indicating the potential value of ontologies to solve the issues of pattern analysis and taking a first step towards a theoretical model for visualization with marine data. It is recommended for future research that marine model developers should intensify their efforts to improve data discovery, sharing, and integration on the base of ontologies.

5. References

[1] K. Malde, N.O. Handegard, L. Eikvil, A.B. Salberg, Machine intelligence and the data-driven future of marine science, ICES J. Mar. Sci. 77 (2020) 1274-1285. doi:10.1093/icesjms/fsz057.

[2] T.R. Gruber, A translation approach to portable ontology specifications, Knowl.Acquis. 5 (1993) 199-220. doi:10.1006/knac.1993.1008.

[3] A. Ferreira, S. Marini, M. Attene, M.J. Fonseca, M. Spagnuolo, J.A. Jorge, B. Falcidieno, Thesaurus-based 3D object retrieval with part-in-whole matching, Int. J. Comput. Vis. 89 (2010) 327-347. doi: 10.1007/s11263-009-0257-6.

[4] M. Zurowietz, T.W. Nattkemper, Unsupervised knowledge transfer for object detection in marine environmental monitoring and exploration, IEEE Access 4 (2020). doi: 10.1109/ACCESS.2020.3014441

[5] B. Kamgar-Parsi, J.L. Jones, A. Rosenfeld, Registration of multiple overlapping range images - scenes without distinctive features, IEEE Trans. Pattern Anal. Mach. Intell. 13 (1991) 857-871. doi: 10.1109/34.93805.

[6] K. Ramar, T.T. Mirnalinee, An ontological representation for Tsunami early warning system, in: Proc. Int. Conf. Adv. Eng., Sci. Manage. (ICAESM), Nagapattinam, Tamil Nadu, India, 30–31 March 2012; IEEE, 2012, pp. 93-98.

[7] R. Lou, Z. Lv, S. Dang, T. Su, X. Li, Application of machine learning in ocean data. Multimed. Syst. (2021). doi:10.1007/s00530-020-00733-x.

[8] N. Boucquey, K.St. Martin, L. Fairbanks, L.M. Campbell, S. Wise, Ocean data portals: performing a new infrastructure for ocean governance, Environ. Plann. D 37 (2019) 484-503. doi: 10.1177/0263775818822829

A complete list of references is available from the author.