A KNOWLEDGE-CENTRIC E-RESEARCH PLATFORM FOR MARINE LIFE AND OCEANOGRAPHIC RESEARCH

Ali Daniyal, Samina Abidi, Ashraf AbuSharekh, Mei Kuan Wong and S. S. R. Abidi
Department of Computer Science, Dalhousie University, 6050 Univeristy Ave, Halifax, Canada

Keywords: e-Research, Knowledge management, Web services, Marine life, Oceanography.

Abstract: In this paper we present a knowledge centric e-Research platform to support collaboration between two diverse scientific communities—i.e. Oceanography and Marine Life domains. The Platform for Ocean Knowledge Management (POKM) offers a services oriented framework to facilitate the sharing, discovery and visualization of multi-modal data and scientific models. To establish interoperability between two diverse domain, we have developed a common OWL-based domain ontology that captures and interrelates concepts from the two different domain. POKM also provide semantic descriptions of the functionalities of a range of e-research oriented web services through a OWL-S service ontology that supports dynamic discovery and invocation of services. POKM has been deployed as a web-based prototype system that is capable of fetching, sharing and visualizing marine animal detection and oceanographic data from multiple global data sources.

1 INTRODUCTION

To comprehensively understand how changes to the ecosystem impact the ocean's physical and biological parameters (Cummings, 2005) oceanographers and marine biologists—termed as the oceanographic research community—are seeking more collaboration in terms of sharing domain-specific data and knowledge (Bos, 2007).

To support the oceanographic research community we have developed a collaborative e-research platform to enable the timely sharing of (i) multi-modal data collected from different geographic sites, (ii) complex simulation models developed by specialized research teams; (iii) high-dimensional simulation results, generated by specialized simulation models, reflecting the local dynamics of specific geographic regions; and (iv) textual knowledge resources—i.e. research articles, reports, news and case studies.

We propose a knowledge management approach to complement observation-based research programs with high-level knowledge-based models to assist researchers in establishing the causal, associative and taxonomic relations between raw data, modelled observations and published knowledge.

We present an e-research platform termed *Platform for Ocean Knowledge Management (POKM)*—that offers a suite of knowledge-centric services for oceanographic researchers to (a) access, share, integrate and operationalize the data, models and knowledge resources available at multiple sites; (b) collaborate in joint scientific research experiments by sharing resources, results, expertise and models; and (c) form a virtual community of researchers, marine resource managers, policy makers and climate change specialists. POKM is supported by the CANARIE network (Canada's high bandwidth network) that enables the rapid collection and integration of high-volume oceanographic data and knowledge (in text format) from distributed sites, and to broadcast the high-dimensional results to users across the world (Barjak, 2006).

The design approach for POKM is to integrate Knowledge Management (KM) technologies with Service Oriented Architectures (SOA). In order to meet the abovementioned functional capabilities of the e-research platform—i.e. POKM—we pursued a high-level abstraction of ocean and marine science domains to establish a high-level conceptual interoperability between the two domains. This is achieved by developing a rich domain ontology that captures concepts from both domains and interrelates them to establish conceptual, terminological and data interoperability. To define the functional aspects of the e-research services we

Daniyal A., Abidi S., AbuSharekh A., Kuan Wong M. and S. R. Abidi S..
A KNOWLEDGE-CENTRIC E-RESEARCH PLATFORM FOR MARINE LIFE AND OCEANOGRAPHIC RESEARCH.
DOI: 10.5220/0003101703630366
In *Proceedings of the International Conference on Knowledge Management and Information Sharing* (KMIS-2010), pages 363-366
ISBN: 978-989-8425-30-0

have developed a services ontology that provides a semantic description of knowledge-centric e-research services. These semantic descriptions of the e-research services are used to both establish correlations between domain and functional concepts that are the basis for data and knowledge sharing, cataloguing and visualization.

In this paper, we present the knowledge management functionalities achieved through the definition of the domain and service ontologies.

2 KNOWLEDGE RESOURCE LAYER

The knowledge resource layer constitutes:

- Data repositories for ocean data, marine life data and simulated data generated through various simulations.
- Domain-specific knowledge represented as research papers and technical reports.
- Domain-specific information represented as images, movies, audio and posters.
- Simulation models shared by the researchers.

2.1 Ontological Modeling of the Domain

The domain ontology in POKM serves as the formal semantic description of the concepts and relationships pertaining to the Marine Biology and the Oceanography domain. POKM provides a core ontology that contains concepts necessary for modelling Marine Animal Detection Data (MADD), Oceanography Data, data transformations and interfaces of the web services in POKM.

The taxonomic hierarchy of the domain ontology constitutes 20 highest level classes.; 15 of these classes are further decomposed into sub-classes at the lower levels of hierarchy.

2.2 Modelling Marine Sciences

There are six upper level classes related to marine sciences—i.e. MARINEORGANISM, ANIMALDETAIL, TAXONOMY, TAXONID, MARINELIFEDATA, MARINELIFEDATACOLLECTION, DATASOURCE, DATAFORMAT.

MARINEORGANISM represents all marine animals, plants and plankton via classes MARINEANIMAL, MARINEPLANT and PLANKTON respectively. There are four main subclasses of MARINEANIMAL: FISH, MARINEMAMMAL, REPTILE and SEABIRD. MARINEPLANT has two main sub-classes: ALGAE and SEAGRASSES. Plankton has three sub-classes representing three functional groups of planktons: BACTERIOPLANKTON, PHYTOPLANKTON and ZOOPLANKTON.

ANIMALDETAIL represents all the necessary information to build a marine animal profile. It has five main sub-classes: AGE, LIFESTAGE, MOVEMENTBEHAVIOR, SEX, TAGID. AGE represents the age of the animal. LIFESTAGE represents the current stage of the animal in life, e.g. adult, juvenile, sub-adult etc. MOVEMENTBEHAVIOR represents various movement behaviors of marine animal that are captured by its sub-classes: BEHAVIORALSWITCHING, DISPERSAL, DIVING, DRIFT, FORAGING, MIGRATING and MOVEMENTPATTERN.

TAXONOMY represents nine main taxonomic ranks used to categorize marine organisms as follows: CLASS, FAMILY, GENUS, KINGDOM, ORDER, PHYLUM, SCIENTIFICNAME, SCIENTIFICNAMEAUTHOR and SPECIES.

TAXONID describes an organism in terms of the above mentioned nine taxonomic ranks.

MARINELIFEDATA represents various aspects of the data about the marine organisms. These include temporal data represented by sub-classes: DAYCOLLECTED, MONTHCOLLECTED, YEARCOLLECTED, DATELASTMODIFIED and TIMESTAMPCOLLECTED, which has two sub-classes of its own: ENDTIMESTAMPCOLLECTED and STARTTIMESTAMPCOLLECTED. The class MARINELIFEDATA is also used to represent concepts related to the cache of the marine data that is represented using sub-classes such as CACHEID, RECORDLASTCACHED, BASISOFRECORD and RESOURCEID.

MARINELIFEDATACOLLECTION is a class the properties of which are used to capture all the data represented by class MARINELIFEDATA.

2.3 Modeling Ocean Sciences

The classes to model ocean sciences include: OCEANREGION, OCEANPARAMETER, SATELLITEINFORMATION, INSTRUMENT, MEASURE, MOVEMENTMODEL, MODELATTRIBUTE, FILETYPE,

OCEANREGION represents all ocean regions categorized by five main sub-classes: ARCTICOCEAN, ATLANTICOCEAN, INDIANOCEAN, PACIFICOCEAN and SOUTHERNOCEAN. Each of these classes are further sub-divided into sub-classes representing sub regions of the each ocean region.

OCEANPARAMETER represents all the geophysical parameters used to describe an ocean environment. These are modeled as sub-classes of this class and include parameters about:

- Air, such as: AirTemperature,
- Wind, such as: WindGust, WindSpeed, EastWardWind, NorthWardWind, UpWardWind and WindFromDirection
- Water, such as: WaterDepth, WaterTemperature, Salinity and Density,
- Current, such as: CurrentToDirection, EastWordCurrent, NorthWardCurrent, FlowVelocity and UpWardCurrent
- Sea Layers, such as: SeaSurfaceElevation, SeaSurfaceTemperature and Thermocline

SatelliteInformation represents the satellite used to monitor the oceans, represented in terms of nine sub-classes: SatelliteID, Altitude, BestSignalStrength, FrequencyOfTransmission, ElapsedTime, NumOfMessagesRecieved, NumOfSuccessfulPlausibleChecks, QualityIndicator and SensorChannel.

Instrument represents all the instruments used for the observation of oceans and to measure various parameters, such as: temperature, salinity and density of the ocean water, ocean currents, depth, pressure, etc. These instruments are represented as the following sub-classes: ADCP, Argos, ArgoFloat, CTD, ElectronicTag, Glider, GlobalPositioningSystem, Satellite and SubmersibleRadiometer.

Measure represents all the spatial and temporal measures of the regions used in the domain of Ocean Sciences, and are modelled as two main sub-classes SpatialMeasure and TemporalMeasure respectively. The sub-class SpatialMeasure has further sub-classes: Height, Latitude, Longitude and SpatialResolution representing the respective spatial measures of the relevant ocean region. TemporalMeasure has two sub-classes: TimeInterval and TimeResolution, representing the respective temporal measures.

MovementModel represents various models used to estimate the migrating and foraging behaviors of marine organisms and their movement parameters such as determining the next positioning estimate of an animal after a period of missing data. These models are represented as sub-classes: FirstPassageTime, FractalAnalysis, GeoLocationModel, KernelAnalysis, StateSpaceModel.

Unit represents all the units used to measure geophysical parameters describing an ocean. It has nine sub-classes: DensityUnit, DepthUnit, LightLevelUnit, SalinityUnit, SpatialResolutionUnit, SpatialUnit, TemperatureUnit, TimeUnit, VelocityUnit.

2.4 Relationships Between Classes

In addition to providing semantics for modelling different resources on the POKM system, the purpose of the domain ontology is to inter-relate the domains of Marine Sciences and Ocean Sciences. There are seventy seven object properties and six datatype properties. We describe only the salient properties are described in this section.

The class MarineAnimal (sub-class of MarineOrganism) is related to respective sub-classes of the class MarineLifeData through properties *has_age, has_sex, has_life_stage, has_movement_behavior* and *has_TagID.* In addition it is also related to class OceanRegion through property *has_geographic_area*. Thus, this property relate the domains of marine sciences and ocean sciences.

The class OceanParameter is related to class Unit through property *has_unit*. This property is given hasValue restriction, to restrict the filler of the property to a specific instance of the class Unit. For example AirTemperature, which is an OceanParameter *has_unit* Degree Celsius, which is an instance to class Unit.

The class MarineLifeDataCollection is related to respective sub-classes of class MarineLifeData through properties *has_basis_of_record, has_cache_ID, has_date _last_modified, has_day_collected, has_depth, has_depth_precision, has_latitude, has_longitude, has_month_collected, has_record_last_cached, has_record_ID, has_taxon_ID, has_temperature, has_time_of_display_collected, has_time_zone_collected* and *has_year_collected.* Each one of these properties is a functional property.

The class MovementModel is related to respective sub-classes of class ModelAttribute through properties: *has_hierarchical, has_input_data, has_linearity, has_observation_error, has_output, has_statistical_ estimation_method, has_statistical_framework, has_stochasticity* and *has_time_value*.

Each OceanRegion is related to various OceanParameters through properties: *has_density, has_flow_velocity, has_salinity, has_sea_surface_elevation, has_water_depth, has_water_mass* and *water_temperature*. Class OceanRegion is also related to respective sub-classes of class MarineLife through sub-classes *has_marine_animal, has_marine_plant* and *has_plankton.* Note that these three sub-classes relate the ocean sciences domain with marine sciences domain.

The class TAXONID is related with respective sub-classes of class TAXONOMY, in order to capture the identification features of each of the marine species. These properties are: *has_class, has_family, has_genus, has_kingdom, has_order, has_phylum, has_sceintific_name, had_scientific_name_author* and *has_species*. Each one of these properties is a functional property.

The class SATELLITE is related to respective sub-classes of SATELLITEINFORMATION through properties: *has_altitude, has_best_signal_strength, has_elapsed_time, has_frequency_of_transmission, has_num_of_messages_received, has_num_of_successful_plausible_check, has_quality_indicator, has_satellite_ID* and *has_sensor_chanel*.

3 MANAGING DATA-SOURCES

POKM provides users the ability to procure Marine Animal Detection Data (MADD) and oceanographic data from multiple sources. Oceanographic data is stored in netCDF format and is accessible by means of standardized methods. However MADD is normally stored in a relational database with no standard relational structure. The relational structure used to store MADD varies from one data source to another. To support this functionality, we have modelled MADD using the domain ontology. Our approach for modelling MADD renders it possible for users to (a) integrate additional MADD sources; and (b) develop a high-level querying mechanism to enable data access from heterogeneous MADD sources.

In order to retrieve MADD from different sources, POKM needs to access different relational structures. We employ the domain ontology to provide that common vocabulary. We define each MADD source as a DATASOURCE in the domain ontology, whereby each DATASOURCE has a number of tables captured by instances of the TABLE class. Each instance of the class TABLE has a number of COLUMNs related to it, which in turn is mapped to a concept in the domain ontology.

In addition to facilitating end-users in writing detailed queries to obtain data from MADD sources. POKM also requires some high level querying mechanisms to enable certain functionalities on the portal. For example to be able to query MADD using just the temporal coverage and bounding boxes (spatial coverage) requires a query specific to the relational structure of the MADD source.

4 CONCLUSIONS

POKM presents a highly distributed e-research and e-science environment that enables researchers distributed across multiple locations are able to collaborate through a suite of services

The oceanographic research community generally uses a wide variety of internet resources (projects, frameworks, systems) in their research. The usual approach, particularly for researchers, has been to download public oceanographic data onto their workstations for integration with privately held biological data, and then to conduct the required analyses. There are vast volumes of oceanographic data repositories, specialized models and data analysis tools that can be leveraged to pursue more comprehensive and complex studies. The efficacy of POKM has been noted by researchers in terms of (a) access to global animal tracking observations, including the animal tracking data provided by OTN; (b) access to a range of models that can be used to interpret, interpolate and extrapolate the animal tracking data; (c) interpretations of animal movement and its causes through evaluation of model uncertainty through a multi-model approach; and (d) most importantly interoperability of data and terminology between the ocean and marine life science communities so that they are now able to collaborate more effectively. POKM users are now able to interconnect distributed raw data, simulation models and knowledge resources into an integrated ‘knowledge asset’ to advance scientific exploration programs.

ACKNOWLEDGEMENTS

This project is supported by a grant from CANARIE (Canada). The authors acknowledge the support of members of the entire POKM team.

REFERENCES

Cummings, J. N., Kiesler, S., 2005. Collaborative research across disciplinary and organizational boundaries. In *Social Studies of Science*, vol. 35(5), pp. 703-722.

Bos, N., Zimmerman, A., Olson, J., Yew, J., Yerkie, J., E. Dahl, 2007. From shared databases to communities of practice: A taxonomy of collaboratories. In *Journal of Computer-Mediated Communication*, vol. 12(2), pp. 652-672.

Barjak, F., 2006. Research productivity in the internet era. In *Scientometrics*, vol. 68(3), pp. 343-360.