Myers, Trina Sharlene (2009) *Applying semantic technologies and artificial intelligence to eco-informatic modelling of coral reef systems.* PhD thesis, James Cook University.

APPLYING SEMANTIC TECHNOLOGIES AND ARTIFICIAL INTELLIGENCE TO ECO-INFORMATIC MODELLING OF CORAL REEF SYSTEMS

Thesis submitted by

Trina Sharlene Myers

November 2009

for the Degree of Doctor of Philosophy

James Cook University

Supervisor:

Professor Ian Atkinson

Statement of Access

I the under-signed, the author of this work, understand that James Cook University will make this thesis available for use within the University Library and, via the Australian Digital Thesis network, for use elsewhere.

I understand that, as an unpublished work, a thesis has significant protection under the Copyright Act and I do not wish to place any restriction on access to this thesis.

________________________ ________________________

Signature Date

Declaration

I declare that this thesis is my own work and has not been submitted in any form for another degree or diploma at any university or other institution of tertiary education. Information derived from the published or unpublished work of others has been acknowledged in the text and a list of references is given.

Signature   Date

Electronic Copy

I the under-signed, the author of this work, declare that the electronic copy of this thesis provided by James Cook University Library is an accurate copy of the print thesis submitted, within the limits of the technology available.

___________________________ ___________________________

Signature Date

Statement on the Contribution of Others

The research described and presented in this thesis was undertaken by the author under supervision by Professor Ian Atkinson and Professor Bill Lavery, both of whom provided editorial and academic advice.

For financial support, I thank Professor Marimuthu Palaniswami of the ARC research network on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP), Department of Electrical and Electronic Engineering from the University of Melbourne for the guidance, support and scholarship in the later stages of the research. I am also grateful to the James Cook University Graduate Research School for an in-kind scholarship and the Faculty of Science and Engineering for two internal Research Grants. I am also appreciative of the travel allowances given by the (then) School of Maths, Physics and Information Technology (JCU) and the DEST funded ARCHER project that made possible the publication and subsequent presentations of this research at both national and international conferences.

I thank Dr. Rosemary Dunn with deep gratitude for her infectious love of the written word. Her editorial advice and tutorage in literature and writing skills exponentially increased the standard of this thesis.

I would like to thank Dr. Ron Johnstone, Dr. Glen Homes and Jeff Maynard for their helpful discussions and expertise in the marine domain.

Acknowledgments

Firstly I would like to thank my husband Michael for being my biggest fan. The journey has been both exciting and arduous and throughout it Michael has been there as a powerful source of motivation and strength with his eternally positive attitude. I would like to thank my parents, Dr. Lance and Beth Myers, for also being incredibly supportive. Their belief in my ability and potential has always been a strong influence in my life and I know, because of this belief, every challenge I undertake will be successfully accomplished.

I owe deep gratitude to my primary supervisor Professor Ian Atkinson who has been an amazing mentor. Ian has never wavered in his belief in me and my abilities, even in times when I questioned it myself. Ian’s support and guidance throughout the journey was superb, he grounds a student and clears their mind in his unique way so the scope is always visible and the goals are always attainable. I would like to thank my secondary supervisor Professor Bill Lavery for his faith in me and his support throughout the journey.

Finally I would like to thank all of those that contributed to this thesis through sheer faith, collaboration, discussion, and/or feedback and with special thanks to: David Browning, Louise Dowling, Dr. Ickjai Lee, Dianna Madden, Associate Professor Richard Monypenny, Nigel Sim and Dr. Jarrod Trevathan.

List of Publications

Myers, T., Atkinson, I. & Johnstone, R. 2010, 'Semantically enabling the SEMAT project: extending marine sensor networks for decision support and hypothesis testing (accepted)', *3rd International Workshop on Ontology Alignment and Visualization (OnAV 10)*, Krakow, Poland, 15 - 18 February, IEEE.

Myers, T. S., Atkinson, I. M. & Johnstone, R. 2009, 'Supporting coral reef ecosystems research through modelling a re-usable ontology framework', *Journal of Applied Artificial Intelligence*, vol. 90, no. 24**,** pp. (in press).

Myers, T. S., Atkinson, I. & Johnstone, R. 2008, 'Supporting coral reef ecosystems research through modelling re-usable ontologies', *Proceedings from the Knowledge Representation Ontology Workshop (KROW 2008)*, Sydney, Australia, 17 September, ACS, pp. 51-59.

Myers, T. S. & Atkinson, I. M. 2008, 'The Semantic Reef: A hypothesis-based, eco-informatics platform to support automated knowledge discovery for remotely monitored reef systems.', *Proceedings of the 11th International Coral Reef Symposium (ICRS 08)*, Ft. Lauderdale, FL, USA, 7-11 July.

Myers, T. S., Atkinson, I. M. & Maynard, J. 2007, 'The Semantic Reef: An eco-informatics approach for modelling coral bleaching within the Great Barrier Reef', *Environmental Research Event (ERE 07)* Cairns, Australia, Environmental Research Event Organising Committee.

Myers, T. S., Atkinson, I. M. & Lavery, W. J. 2007, 'The Semantic Reef: Managing complex knowledge to predict coral bleaching on the Great Barrier Reef', *Proceedings of the fifth Australasian symposium on ACSW frontiers*, ACS, Ballarat, Australia, vol 68, pp. 59-67.

In loving memory of my beautiful little angel

Tameka

Abstract

A "data deluge" is overwhelming many areas of research. Massive amounts of scientific data are being produced that cannot be effectively processed. Remote environmental monitoring (including sensor networks) is being rapidly developed and adopted for collecting real-time data across widely distributed locations. As the volume of raw data increases, it is envisaged that bottlenecks will develop in the data analysis phase of research workflows, because data processing and synthesis procedures still generally involve manual manipulation.

Despite the exponential growth in data and the consequential challenges in data management, current e-Research communities are exploring solutions to the "data deluge". E-Research is the amalgamation of research techniques, data and people with Information Communication Technologies (ICT) to enhance research capabilities. Recent research efforts by the Semantic Web and Knowledge Representation (KR) domains focus on the development of automated data synthesis technologies. A key component in these solutions is the semantic technologies. Semantic technologies involve methods to add contextual information to data through ontologies so logic systems can be applied by the computer to enable automated inference. An ontology explicitly describes concepts in "computer-understandable" terms which allows for automated reasoning and intelligent decision-making by the machine. Automated data analysis and knowledge discovery is desirable because the manual manipulation of data processing and synthesis requires human intervention which will become increasingly more difficult to sustain as the data deluge grows.

This dissertation introduces the Semantic Reef project which is an eco-informatics software architecture designed to alleviate data management problems within marine research. The intention was to develop an automated data processing, problem-solving and knowledge discovery system within the scope of e-Research, which will assist in developing our understanding and management of coral reef ecosystems. The Semantic Reef project employs e-Research approaches including semantic technologies and scientific workflows, which together create a platform designed to evaluate complex hypothesis queries and/or provide alerting for unusual events (e.g., coral spawning or bleaching).

The Semantic Reef project was built as a KR platform, so researchers can combine disjoint data from different sources into a single Knowledge Base (KB) to pose questions of the data. Scientific workflows access and retrieve remote sensor data and/or data available via the Web to

populate the KB. The KB consists of a hierarchy of reusable and usable ontologies that together generically model a coral reef ecosystem in a "computer-understandable" form. The ontologies range from informal through to formal and, when coupled to datasets, derive inferences from data to "ask" the KB questions for semantic correlation, synthesis and analysis. The ontology design leverages the scalable and autonomic characteristics of semantic technologies such as modularity, reuse and the ability to link latent connections in data through complex logic systems.

The overall goal of the Semantic Reef project was to enable marine researchers to pose hypotheses about environmental data gathered from *in situ* observations, and to explore phenomena such as climate change effects on an ecosystem rather than on one component at a time. Currently, in marine research, there has been an explosive increase in the number of questions posed about climate change effects; for example, questions about the origins of phenomena such as coral bleaching on coral reef ecosystems. To be answered, these questions need to be able to assess the cumulative combination of ecological factors and stressors that contribute to the tipping point from a healthy coral to stressed coral due to coral bleaching. The marine biology domain has an urgent need for more efficient investigation of the disparate data streams and data sources. The Semantic Reef project, which incorporates the new hypothesis-driven research tools and problem-solving methods, is designed as a proof of concept to resolve this need.

The Semantic Reef system has the capacity to pose hypotheses and automate inferences of the available data. The system's design supports flexibility in theoretic hypothesis design because the researcher is not required to predetermine the exact hypothesis prior to gathering data for import to the KB. Rather, the questions can be as flexible as the researcher requires, and they may evolve as new data becomes available or as ideas grow and/or epiphanies emerge. Then, once phenomena in the data are disclosed through semantic inference, *in situ* observations can be performed to confirm or negate the theory. The Semantic Reef tool offers marine researchers this flexibility in hypothesis modelling to theorise about a range of scientific conundrums such as the cumulative causal factors that contribute to coral bleaching.

This study is the first known example of Semantic Web technologies and scientific workflows combined to integrate data, with the purpose of posing observational hypotheses or inferring alerts in the coral reef domain. As a proof of concept, the Semantic Reef system offers a different approach to the development and execution of observational hypotheses on coral reefs. The system offers adaptability when applying hypotheses and questions of data, specifically in scenarios where the hypothesis is not apparent prior to data collection efforts. The Semantic Reef system cannot overcome the data deluge, but it offers a unique approach to the discovery of new phenomena that, through automation, can alleviate the problems associated with the data analysis phase.

Table of Contents

List of Tables

List of Figures

Glossary of Acronyms

ACRONYM	MEANING
ABS	Australian Bureau of Statistics
AIMS	Australian Institute of Marine Science
API	Application Program Interface
AVHRR	Advanced Very High Resolution Radiometer
AWS	Automatic Weather Station
BOM	Australian Bureau of Meteorology
CC	Creative Commons
CHAMP	Coral Health and Monitoring Program
CRC	Cooperative Research Centre
CREON	Coral Reef Environmental Observatory Network
CSIRO	Commonwealth Scientific and Industrial Research Organisation
CWA	Closed World Assumption
DAML + OIL	Darpa Agent Markup Language plus the European Ontology Interchange Language
DHD	Degree Heating Day
DIG	Description Logic Implementation Group
DL	Description Logics
DLP	Description Logic Programming
DOGMA	Developing Ontology-Grounded Methods and Applications
FOL	First Order Logic
GBIF	Global Biodiversity Information Facility
GBR	Great Barrier Reef

ACRONYM	MEANING
GBRMPA	Great Barrier Reef Marine Park Authority
GBROOS	Great Barrier Reef Ocean Observing System
GEON	GEOscience Network
GIS	Geographic Information System
GLEON	Global Lake Ecological Observatory Network
HCI	Human Computer Interface
HPC	High Performance Computing
HTML	Hypertext Markup Language
HTTP	Hypertext Transfer Protocol
ICON	Integrated Coral Observing Network
IMOS	Integrated Marine Observing System
IPCC	Inter-governmental Panel on Climate Change
ICT	Information Communication Technologies
ITIS	Interagency Taxonomic Information System
JCU	James Cook University
JISC	Joint Information Systems Committee
JRE	Java Runtime Environment
JVM	Java Virtual Machine
KB	Knowledge Base
KR	Knowledge Representation
LHC	Large Hadrons Collider
LMSM	Local Mean Summer Maximum

ACRONYM	MEANING
LMST	Long-term Mean Sea Surface Temperature
LTMP	Long Term Monitoring Program
MMI	Marine Metadata Interoperability
MMM	Maximum Monthly Mean
MPL	Mozilla Public License
NAF	Negation as Failure
NEON	National Ecological Observatory Network
NEPTUNE	North-East Pacific Time-series Undersea Networked Experiments
NESDIS	National Environmental Satellite, Data and Information Service
NOAA	National Oceanic and Atmospheric Administration
NSF	National Science Foundation
OGC	Open Geospatial Consortium
OGSA	Open Grid Services Architecture
ONC	Ocean Networks Canada
OSG	Open Science Grid
OWA	Open World Assumption
OWL	Web Ontology Language
PAR	Photosynthetically Active Radiation
PROWL	Probabilistic Web Ontology Language (OWL)
RDF	Resource Description Framework
RDFS	RDF Schema
RIF	Rule Interchange Format

ACRONYM	MEANING
SBA	Satellite Bleach Alert
SEEK	Science Environment for Ecological Knowledge
SME	Subject Matter Expert
SIOC	Semantically-Interlinked Online Communities
SPARQL	SPARQL Protocol and RDF Query Language
SQWRL	Semantic Query-Enhanced Web Rule Language
SST	Sea Surface Temperature
SST+	SST anomaly
SSW	Semantic Sensor Web
SW	Semantic Web
SWEET	Semantic Web for Earth and Environmental Terminology
SWRL	Semantic Web Rules Language
uBio	Universal Biological Indexer and Organiser
UNA	Unique Name Assumption
URI	Unified Resource Identifiers
URL	Uniform Resource Locator
URN	Uniform Resource Name
VO	Virtual Organisations
VRE	Virtual Research Environment
W3C	World Wide Web Consortium
WWW	World Wide Web
XML	eXtensible Markup Language