



## Smart Information Retrieval for Marine Sensor Data

Md Sumon Shahriar, Paulo De Souza, and Greg Timms

Tasmanian ICT Centre, CSIRO, Hobart, Australia (mdsumon.shahriar@csiro.au)

Data management and information retrieval (query/search) [1, 2] are regarded as challenging tasks for sensor networks. Specifically for marine sensor networks that include sensors for different phenomena (e.g., water temperature, pressure, and dissolved oxygen), information retrieval needs to consider both streaming and historical data, diverse data formats and storages, data frequency, data quality and the volume of data. We propose a smart (intelligent) system for information retrieval augmenting data mining techniques and semantics in marine sensor networks.

We consider the following smart information retrieval techniques for marine sensor data.

**Data publishing and event detection using continuous queries:** Publishing relevant information [3] and detecting events/patterns [4] both for real-time streaming and historical data using continuous queries are necessary for situation awareness and decision making.

**Pattern query with knowledge discovery:** Pattern queries find similar patterns in marine sensor data. The pattern queries can be implemented both for streaming data and historical data. The pattern query will not only give similar patterns, but also show whether the patterns have increasing or decreasing trend and seasonality using knowledge discovery.

**User, context and query profiling:** We aim to consider different users of marine sensor data and the context of using marine sensor data to profile queries. By query profiling, we can identify different types of queries such as continuous queries and snap shot or historical queries.

**Query intention and relaxation:** Based on the users, context and query types, query intention can be analysed and the query may be relaxed for finding hidden interesting patterns in marine sensor data.

**Semantics in the marine sensor web:** With the emergence of the semantic sensor web, there is a need to extend query processing on the marine sensor web to the marine semantic sensor web (MSSW). The concept of distributed ontologies can be used in query processing over MSSW. Queries using semantic reasoning techniques [5] can also be developed over the MSSW. We propose to consider: (a) adding ontologies to the marine sensor web, (b) reasoning over the MSSW and (c) distributed query processing over the MSSW.

We consider the following steps for improving marine sensor data processing towards smart information retrieval (query processing).

**Data quality and data cleaning:** Quality marine sensor data are necessary to get quality answers from query processing. Missing or erroneous data affect query processing and answering. Different data mining techniques such as outlier detection need to be considered in measuring data quality.

**Data transformation and integration:** As marine data are represented in a number of formats including model output in various forms of netCDF [6] and historical sensor data in relational databases, data need to be transformed and integrated for query processing. Data obtained from disparate sensor networks in various domains (e.g., hydrology) also need to be integrated for query processing.

**Statistical processing and clustering:** As raw marine sensor data contain mainly spatial and temporal data, the answer to queries may not have an exact result because data may not exist in the specified time and space. Thus, there is a need to use statistical methods such as interpolation and extrapolation for gap filling or prediction. Marine sensor time series in nature may exhibit seasonal properties. Thus we propose to cluster marine sensor

data using some clustering techniques considering seasonality for efficient query processing.

Pattern discovery: In marine sensor data, there may be no exact answer for a query asking for interesting patterns or behaviours. In this case, data mining techniques such as pattern extraction (e.g., Dynamic Time Warping, ARIMA) and machine learning (e.g., Self-Organizing Map) may need to be exploited.

The proposed smart information retrieval system will benefit industries and sectors that require knowledge from the marine sensor domain for decision making, planning and management.

#### References

- [1] K. Aberer J.Wu, Y.Zhou and K.Tan, "Towards integrated and efficient scientific sensor data processing: A database approach", EDBT, pp. 922–933, 2009.
- [2] G.Kollios M.Hadjieleftheriou Y.Tao N.Mamoulis, H.Cao and D.W.Cheung, "Mining, indexing, and querying historical spatiotemporal data", KDD, pp. 236–245, 2004.
- [3] Tarboton D.G. Piasecki M. Maidment D.R. Zaslavsky I. Horsburgh, J.S., "An integrated system for publishing environmental observation data", Environmental Modelling and Software, 24, pp. 879–888, 2009.
- [4] Minsker B.S. Hill, D.J., "Anomaly detection in streaming environmental sensor data: A data-driven approach", Environmental Modelling and Software, 25, pp. 1014–1022, 2010.
- [5] Morris R.A. Peri F. Calder, M., "Machine reasoning about anomalous sensor data", Ecological Informatics, 9, pp. 9–18, 2010.
- [6] "<http://www.unidata.ucar.edu/software/netcdf/>".