2011 International Conference on Environmental and Computer Science
IPCBEE vol.19(2011) © (2011) IACSIT Press, Singapore

Framework for Tropical Ecological Data Warehouse (TEDW) FOR Governance and Maintenance of Tropical Lakes

Hui Cham[1], Sorayya Bibi Malek Binti Malek Abd Rashid[2], Sharifah Mumtazah Syed Ahmad and Aishah Binti Salleh

Abstract. The current data collection on tropical lakes and wetlands especially in Malaysia is carried out in a disparate manner which does facilitate database integration and data sharing. This can be considered as an inefficient use of available resources. Such an issue is partly due to the lack of proper standards on tropical data warehousing that would otherwise provide the best practices for common platform and coherent efforts in lake managements. Therefore the main objective of this paper is to propose a suitable framework for tropical ecological data warehouse (TEDW) of lake and wetlands water bodies. The design of TEDW is robust as it caters for different ecological objects, formats and variables which are desirable for actual data warehouse implementation. In addition, TEDW also provides a real-time data entry with web interfaces as well as basic statistical analysis and graphical representations. In order to facilitate data sharing with outside parties, TEDW provides a restricted access to authenticated parties. This paper discusses the general scope, data structures and implementation of the TEDW framework with the aim of facilitating the best practices in lake management through monitoring and evaluation of ecological data of tropical water bodies.

Keywords: ecological, data warehouse, data mining, web services

1. Introduction

Jusoh [1] reported that 38% and 62% of the 90 monitored lakes in Malaysia were classified as mesotrophic and eutrophic respectively. To prevent eutrophication, the quality of water source needs to be observed and monitored. Hence proper management of ecological data is imperative for effective lake and wetlands monitoring and governance. However, there are no comprehensive monitoring programs carried out in most lakes in Malaysia, except Putrajaya Lakes [2]. In addition, there is no reported study on an integrated approach for proper database development at a national level [3].

At the international level there are many lake databases developed and maintained such as the World Lake Database [4], Ramsar Site Database [5], and GIS WORLDLAKE database [6]. The World Lake Database consist data on 500 lakes from 73 countries. The data is separated into a few categories which are location, description, physical dimensions, lake water quality, physiographic features, biological features, Socio-economic conditions, lake utilization, deterioration of lake environments and hazards, wastewater treatment, improvement works in the lake, legislative and institutional measures for upgrading lake environments, development plans, and sources of data.

The Ramsar Sites Database meanwhile is an internet accessible database which can generate report for public usage. The database includes the details from the Ramsar Information Sheet, the Ramsar National Report and information provided by contract parties. GIS WORLDLAKE consists of data of geographic, morphometric, hydrological, meterological and climatological, hydrochemical, hydrothermal, and others.

However the information contained in these databases on tropical lakes in Malaysia are not comprehensive enough.

+ Corresponding author. Tel.: + 60127609992.
E-mail address: chamhui@siswa.um.edu.my.

Current ecological databases on tropical lakes in Malaysia are scatted as they are compiled and maintained by independent bodies and organizations whereby there are limited evidences that the databases conform to any guidelines. The above mentioned factors do not facilitate database integration or data sharing purposes, which is an inefficient usage of resources. This can be solved by having a database for data archive and integration. However the implementation of ecological database is not straight forward and requires expertise in database programming technology which may not be possessed by biologists or ecologists [7]. Current ecological databases are structured in heterogeneous formats using different platforms. Thus, data sharing between these databases can be a difficult task due to the incompatibility issue.

The aim of this paper is to provide possible solutions for the issues mentioned above by providing an ecological data warehouse (TEDW) with interactive user interface that will reduce the learning curve to user of the database. The TEDW system provide standard for data archiving and retrieval to allow seamlessly data transfer between researchers independently despiteof the database platform used. The TEDW is also envisaged with the capability of proper data formatting capabilities as such the data can be manipulated by other data mining tools. It addition, the system is designed carefully to ensure compatibility with a common Geographical Information System (GIS) tool and data mining system.

2. Methods

2.1. System Architecture

The Tropical Ecological Data Warehouse (TEDW) was built for data management on tropical ecological data. The target users includes both expert and non-experts groups. Thus the system should be user friendly and provide necessary supports suitable for these targeted users. The TEDW in this study can be use as an information management system for ecological data with interactive user interface, capable of generating meaningful statistical report for analysis purposes and formation of suitable files for manipulation of data mining tools. In addition, the TEDW provides necessary support for data exchange and sharing within expert and non-expert groups in related fields with data update capability in a geographical information system (GIS).

Figure 1 represents the process flow of the TEDW. MSSQL is utilized as data acquisition component to provide an interface for loading and cleaning historical or on-line ecological data. Database administrator may insert data from excel, access or text file into database file with default functions of MSSQL. On top of that, ASP.Net platform is utilized as data archiving component and is closely linked to both XML metadata processing as well as data analysis and modelling. TEDW provides a user-friendly interface for users to retrieve information from the database and also generate necessary reports and data files. The users also can send formatted data from TEDW to data mining system to recognize patterns from complex data sets by combining statistical analysis and artificial intelligence techniques.

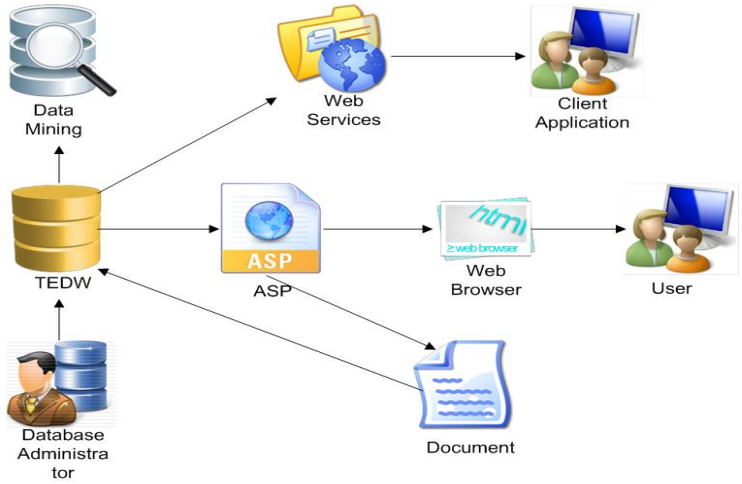

Figure 1

Fig. 1: Process Flow of TEDW

2.2. Data Mining & GIS

The TEDW system has been designed as such data mining techniques can be deployed efficiently. The data mining techniques used by TEDW are (GIS), Kohonen Self Organizing Feature Map (SOM), Artificial Neural Network (ANN), and Hybrid Evolutionary Algorithm (HEA). Machine learning techniques such as SOM, ANN and HEA have shown promising results for tropical lake. SOM has been used in ecological modelling at tropical lake to find similarity between dataset which have yield a reasonable level of accuracy [8]. ANN learn about a particular subject from the data provided, they are trained, rather than being programmed by the users. ANN has been successfully applied for predicting phytoplankton abundance at tropical lake [9]. Meanwhile HEA is able to generate rule sets or arithmetic functions for ecological processes for provides explanation and forecasting for specific output variable such as habitat or water quality conditions. Rules generated from HEA for tropical lake was integrated with thematic map visualization system for prediction of phytoplankton biomass [10].

TEDW system allows interoperability between these machine learning techniques by ensuring data standards compatibility. This is done by providing a module that allows data transformation to suitable format for the machine learning techniques. This module is designed with compatibility issues in mind such as the data can be formatted into appropriate format and exported for the selected data mining technique.

GIS is useful in ecological modelling as it allows clustering, visualization, and discovery of relationship of ecological data [11][12]. TEDW incorporates the functionality of GIS for visualization of selected ecological parameter distribution over region, time frame, and location. This enables discovery of relationship within the ecological dataset that can be provides useful information for lake monitoring and governance.

2.3. Web Services & Standard

Currently GBIF uses Darwin Core [13] standard to develop structured data models and controlled vocabularies in key areas of biodiversity informatics. TEDW adopts Darwin Core standard which has been used by Global Biodiversity Information Facility (GBIF) in web services for information sharing and acquisition [13]. The Darwin Core Standard in TEDW is able to provide high accessibility and flexibility to allow user to access TEDW from other software or systems. In a study carried out by Benjamin [14] technology of web services in geospatial, scientific workflows and related issues are helpful in maintaining biogeographical archive and habitat analysis. Most of the ecological and biological data management systems are using Darwin Core for their data standard and exchange protocol. This allows the users of TEDW to be able to perform data sharing with huge amount of ecological data management system.

TEDW provides web service as a suitable bridge to connect to other systems within similiar fields. Papaxoglou [15], stated that any piece of code or the application component deployed in system can be transformed into a network service. TEDW web services provide a platform for users to discover and invoke this network services based on the idea of composing applications using the standard protocol known as SOAP. Users can program their system to act as a client application and command functions provided in the web services. As a result, TEDW will act as a service provider for expert and non-experts users to integrate the web services functions into their working environments.

Chad[16] reported that complex ecological data collected by researcher are used is many types of protocol to share within their community or even globally. Therefore heterogeneous data should be stored in autonomous database which can be dispersed throughout the ecological research community. TEDW uses Extensible Markup Language (XML) metadata which allows a higher degree of interoperability among distributed research groups [17]. Much research particularly in ecology research[18] has indicate that most of the current ecosystem informatics research use XML to represent metadata.

In addition, TEDW adopts the recommendations from Ramsar Classification System produced by the Wetlands International body. The classification provides a list of wetlands which can assists in rapid identification of main wetland habitats represented from each site.

3. Discussion and Result

The user interface of TEDW comprises of data searching (integrated and simple search), search results, data input (for ecological data and digitized map) and statistical report. TEDW system is developed using ASP.Net. ASP.Net is chosen because it enables real time data entry and access via internet browser. Data entered into TEDW comprises of temporal and spatial environmental data of lake and wetlands using Ramsar classification [18].The information entered into TEDW will be used for management and governance for tropical lake. Search function in TEDW can conduct an integrated search using simple input such as wildcards and selection list for user convenience. The detailed search results are presented by hydrological, biological, chemical, geographical data for specific variable, time and location. The standardization of information is important to prevent incorrect data being from entered into the TEDW. However TEDW allows user to manage environmental variables and quantifiers. This means that the users are allowed to modify it according to the nature of their data.

The statistical and report generation in TEDW allows information to be displayed based on user selection of time, region, location and input variables. Statistical reports and analyzed data can help users to monitor the changes in tropical lakes to prevent eutrophication. The GIS demonstrates data exported from TEDW. The exported data is compatible with ArcGIS and machine learning techniques deployed in this study. Prediction and modelling of environmental data in tropical lakes is essential to allows users to foresee future changes and take action to reduce threats of ecosystems.

4. Conclusion

TEDW demonstrated a high flexibility design of a lake data warehouse which can be adapted to any kind of lake. It provides an organized framework to collect fragmented ecological data and archive it into a standardized method. With internet connectivity, users will be able to access to the TEDW easily. Additionally, part of the functions in TEDW can be integrated into other system as metadata which are well prepared for data migration to be used by other system. The future work for TEDW is focused on computation and modelling on loss of data due to most of the ecological data are collected and compiled from different sources. As a result, data inconsistency frequently occurs in researchers' data sample. A higher accuracy interpolation algorithm or system is required for data mining and visualization.

5. Acknowledgements

This research was supported and funded by University Malaya grant UMRG RG005/09AFR.

6. References

[1] Juhaimi Hj J: Management of Lakes and Reservoirs in Malaysia: An Introduction. Pusat Kajian Kualiti Air dan Alam Sekitar; 2009.

[2] Zati S, and Salmah Z: Lakes and Reservoir in Malaysia: Management and Research Challenges. *In: TAAL 2007: The 12 World Lake Conference*, pp 1349-55. Ministry of Environment & Forests, Govt of India, Jaipur, India.

[3] NAHRIM: A Desktop Study on the Status of Lake Eutrophication in Malaysia. Final Report, 2005.

[4] International Lake Environment Committee Foundation (ILEC): World Lake Database http://wldb.ilec.or.jp/

[5] Wetlands International: Ramsar Sites Information Service http://ramsar.wetlands.org/

[6] Kirill Ya. Kondratyev and N. N. Filatov: Limnology and Remote Sensing, Spinrger, 1999

[7] Judith B.C , Nalini N, Michael F, Anne F, Emerson MH, Lois D, and David M: Component-based end-user database design for ecologists. *J Intell Inf Syst* 2007, 29:7-24

[8] Sorayya M, Aishah S, and Sharifah MSA: Analysis of Algal Growth Using Kohonen Self Organizing Feature Map (SOM) and its Prediction Using Rule Based Expert System. *Proceeding ICIME '09 Proceedings of the 2009 International Conference on Information Management and Engineering. IEEE Computer Society*, Washington DC; 2009.

[9] Oh JH, Sorayya M, Aishah S, and Mohd SB: A Comparison of ANN Architecture towards Predicting Cyanobacteria Abundance at Tropical Putrajaya Lake and Wetlands. *International Conference on Environmental Science and Technology ICEST* 2010.

[10] Lau CF, Sorayya M, and Aishah S: Algal Growth Prediction Using Hybrid Evolutionary Algorithm and Visualization Using Thematic Map Technology. *International Conference Remote Sensing*; 2010.

[11] 10. Xavier S, Jose CB, Neftali S, Juan MP, Gustavo AL, Soumia F, and Xavier P: Inferring habitat-suitability areas with ecological modeling techniques and GIS: A contribution to assess the conservation status of Vipera Latastei. *Biol Cons* 2006, 416-425

[12] Richard A, and Diane P: Integrated geographical assessment of environmental condition in water catchments: Linking landscape ecology, environmental modeling and GIS. *J Environ Manage* 2000, 59:299-319.

[13] GBIF – Global Biodiversity Information Facility. http://www.gbif.org/

[14] Benjamin DB, Patrick NH, Ei F, Andrew JR, Song SQ, Lucie JH, and Robert SS: Geospatial web services within a scientific workflow: Predicting marine mammal habitats in a dynamic environment. *Ecol Inf* 2007. 2(3): 210-223.

[15] Michael PP, Paolo T, Schahram D, and Frank L: Service-Oriented Computing: A Research Roadmap. *Int J Coop Inf Sys 2008*, 17(2) :223-255.

[16] Chad B, Matthew J, Jivka B, and Daniel H: Metacat: a Schema-Independent XML Database System. National Center for Ecological Analysis and Synthesis (NCEAS); 2001

[17] Seligman L, and Roenthal: XML's impact an databases and data sharing. Computer and Processing 2001, IEEE Computer Society. 34: 59-67.

[18] P Vos, E Meelis, and WJ Ter Keurs: A Framework for the Design of Ecological Monitoring Programs as a Tool for Environmental and Nature Management. Leiden University, Netherlands; 1999.