Integrated Database Construction for Efficient Support of Yeongsan River Estuary Management System

G H Lee1, K H Kim2, S J Lee3
Department of Geoinformatic Engineering, Inha University

E-mail:oiuy@inha.edu

Abstract. Yeongsan River is one of the four major rivers in South Korea, and it flows toward the Yellow Sea by passing through Damyang and Gwangju. In particular, the skewness of the main stream in Yeongsan River is relatively higher compared to other rivers. Accordingly, flood damage occurred frequently due to the flooding of sea water during tidal periods. Additionally, the environment of the estuary in Yeongsan River has been severely damaged due to indiscreet development and the inflow of various waste waters. Therefore, water quality improvement and management are crucial. For better water quality management, the government ministry is collecting various data from different fields to identify the water quality conditions. The necessity of collected data is being heightened in order to apply them into the estuary management system. However, in terms of the observed data, the observed field or items frequently modified according to social interests. Additionally, index is needed in order to search for massive amount of observation data. Due to this, the process of construction into database is relatively difficult. Therefore, in this study, these characteristics were considered for construction into the integrated DB.

1. Introduction

An estuary refers to the place where a river flows into ocean. There are two types of estuary such as open estuaries and closed estuaries. The open estuaries have no artificial structures so that sea water and freshwater can be circulated naturally to be purified. However, the closed estuaries have certain types of artificial structures to prevent damages such as flooding or seawater inflow. Accordingly, the circulation of sea water and freshwater is constantly suspended in the closed estuary. For this reason, the natural purification function does not operate as well. Recently, there have been growth concerns regarding the restoration of closed estuaries. Some of the methods include seawater circulation and wetland restoration. Besides this, numerous R&D projects are in progress by the government ministry to identify the current conditions of the estuaries. For this reason, a massive amount of data is being collected from various fields related to estuaries. Furthermore, collected data are increasing in application as the base data for estuary management systems. In previous study, [1] and others conducted a study regarding the construction and application of ecological and environment integrated DB in Yeongsan River estuaries based on GIS. However, in the precedent study, DB was designed by only applying observed data from 2010. Therefore, the modification of the observed fields and items were not considered during the design process. Eventually, the tables have to be newly created every time observed data is collected each year. This contains the problems of memory waste as well as causing the search function to become complex. Other than this, there is an absence of index in DB, which is a crucial factor in the search speed. In case of a dynamic data or sedimentary data, it takes a lot much time to search for them due to the significant amount of data. Accordingly, an
index is an essential condition. Therefore, in this study, the structure of the DB was improved and index was created in order to solve the problems of integrated DB mentioned in the precedent study. Additionally, the observed data collected additionally in 2012 were added into the newly created DB.

2. Study contents and methods

2.1. Current estuary data

The purpose of this study was to analyze the problems of the constructed integrated DB and make improvements. In the preceding study, only data observed in 2010 were used to design the DB. In this study, the data observed during 2010 to 2012 were collected additionally. Generally, the data are classified between the fields of water quality, ecology, and dynamic sediment. Table 1 is a representation of the annual detailed fields, observed themes, and observed items. In the observed item, items (date, station, water level, time, depth) related to the search of observed data and measured items (temperature, salinity, pH, DO) which represent the actual observed value are presented.

Table 1. Collected observation data.

| Category          | Year | Detail field          | Observed themes          | Observed items                                                                 | Number of record |
|-------------------|------|-----------------------|--------------------------|-------------------------------------------------------------------------------|------------------|
| **Water Quality** |      |                       |                          |                                                                               |                  |
|                   | 2010 | Estuary Sediment layer| Water Quality            | Date, time, station, water level, depth, temperature, salinity, pH, DO, NO₃, COD | 78               |
|                   |      | Estuary Lake Sediment layer | Heavy metal, TOC, Grading | Date, station, Al, Fe, As, Cd, Cr, Cu, Mn                                   | 29               |
|                   |      | Estuary Lake Sediment layer | Water Quality          | Date, station, depth, water level, COD, SS, TOC, POC, DOC, TN, PON, DON      | 92               |
|                   | 2011 | Estuary Sediment layer | Surface Sediment, Denitrification | Date, station gra, sand, silt, clay,                                          | 6905             |
|                   |      | Estuary Lake Sediment layer | Water Quality          | date, station, time, depth, temperature, salinity, pH                       | 27               |
|                   | 2012 | Estuary               | Water Quality            | Date, station, water level, depth, temperature, salinity, pH, DO, SS, COD, NH | 196              |
|                   |      | River · Estuary Lake | Water Quality            | date, station, time, depth, temperature, salinity, pH, DO, SS, COD, NH       | 124              |
| **Biological**    |      |                       |                          |                                                                               |                  |
| Phytoplankton     | 2010 | Species, Population | Species, Population      | Date, station, water level, whole chl-a, Net chl-a, nano chl-a               | 457              |
| Zooplankton       |      | Species, Population | Species, Population      | Date, station, time, total abundance, total species, Fresh abundance         | 44               |
| Benthic ecosystem |      | Species, Population | Species, Population, Class | Date, station, species, abundance, class                                    | 597              |
| Phytoplankton     | 2011 | Biomass, Species composition |                      | Date, station, water level, whole chl-a, micro chl-a, nano chl-a, pico chl-a | 1695             |
| Zooplankton       |      | Species, Population | Species, Population      | Date, station, water level, zooplankton carbon biomass, abundance            | 204              |
| Benthic           |      | Species               | Species                  | Date, station, species, abundance                                            | 1393             |
### Existing DB problem

In the precedent study, each of the observed items was formed into a single field in DB table. Therefore, in order to reflect the changes made in the observed items to the DB, methods of 1) creating a new field or 2) newly creating tables like the precedent study can be applied. When a new field is created, there is no value that can be inserted in the previous field and this leads to the problem of being filled with a null value. This signifies the waste of memory existing in null. In particular, in case of data with large quantities of observed amounts, the null value may increase in sum. Therefore, the 1) method is inappropriate. In the 2) method, the number of tables increase greatly and become complicated as the observed item changes. This brings about another problem of waste in memory. Furthermore, it becomes bothersome since SQL must be created differently per table. When considering the fact that various data will continuously enter in the future the 2) method is also

| Ecosystem | Class | Date, station, water level, bacteria production, biomass, reduction | 125 |
|-----------|-------|---------------------------------------------------------------|-----|
| Benthic ecosystem | Species, Population, Biomass | Date, station, water level, species, abundance, biomass | 44 |
| Phytoplankton | Chlorophyll-a, Species composition | Date, station, whole chl-a, micro chl-a, nano chl-a, pico chl-a, Kd, Temperature | 716 |
| Zooplankton | Biomass, Death rate | Date, station, water level, zooplankton carbon biomass, abundance | 154 |
| Microbe - Bacteria | Bacteria Production, Total oxygen, Nutrient flux | Date, station, bacteria production, SRR, TOU | 150 |
| Nutrient cycle | DO, Temperature, Salinity | Date, station, water level, SOD, WOD | 15093 |
| Dynamic Sediment | Temperature, Salinity, Turbidity | Date, station, time, depth, SSC, velocity, direction, temperature, salinity, turbidity | 6246 |
| Water Dynamic | Velocity, Direction | Date, station, depth, dist, direction, velocity | 1193 |
| Sediment1 | ADCP, CTD, SonarAlt | Date, station, time, depth, speed, direction, salinity, SSC, | 32454 |
| Sediment2 | SSC, SS-flux | Date, station, depth, water level, direction, dist, depth | 29177 |
| Dynamic Sediment | Velocity, Direction | Date, station, depth, temperature, salinity, DO, velocity, direction | 91553 |
| Water Dynamic | Salinity, Temperature, Turbidity | Date, station, depth, current direction, current speed, signal strength, salinity, temperature, density | 105915 |
| Sediment | Salinity, Temperature, Turbidity | Date, station, particle construction, sediment, organic matter, carbonate | 142 |
inappropriate. Besides this, index was not created in the precedent study. This can also become a fatal problem when considering the amount of data that will be accumulated in the future.

2.3. Improved DB building
In this study, the DB structure was redesigned in order to prevent the DB from becoming influenced even if the observed item became modified. The tables were distinguished between the data table where actually observed data value are saved and the description table where specifications about tables are saved. First, it is content about the design process of the data table. Tables were created for each detail field since the observed items showed significant differences following the different detail fields. Additionally, the field name of each data table was formed like 'search field1', 'search field2', 'observe field1', 'observe field2', and so on. Accordingly, there is no data in the table that needs to be modified. However, the 'search field1' and 'observe field1' cannot be identified. Therefore, what the 'search field1' and 'observe field1' must be specified. The description table contains specified descriptions.

Table 2. Modified contents.

| Classification | Existing DB | Improved DB |
|----------------|-------------|-------------|
| Table name     |             |             |
| Ex) Estuary_Waterquality10 | Estuary_Waterquality |
| Estuary_Waterquality11 | Estuary_Waterquality_description |
| Estuary_Waterquality12 |             |
| Table classification | non classification | Data table Description table |
| Field name     |             |             |
| Date, station, depth, temperature, TN, BOD, … | search field1, search field2,… |
| observe field1 | observe field2,… |
| Index          | X           | O           |

Next, an index was formed in each data table. In case of the index, each record must be classified. For the observed data used in this study, all data can be distinguished. There is sufficient data for the classification of observation year, observation date, observation station, observation time, water layer and depth. In addition, there is no case such as measuring temperature twice at the same date, same station, and same depth. Accordingly, a composite key was created by synthesizing the search field and applied as an index. In order to define the index for each record, an index design chart was created for each table. Table 2 presents an index design chart of 'estuary water quality'. For instance, the index of the observation value became '1010602YS10930B' in the bottom of the YS1 on 09:30 of June 2nd, 2010 by continuous observation.

Table 3. Example of index design chart (Estuary Water Quality).

| Search Field | Estuary Water Quality Field Value | Index |
|--------------|----------------------------------|-------|
| Year         |                                  |       |
| 2010         |                                  | 10    |
| 2011         |                                  | 11    |
| 2012         |                                  | 12    |
| Basic        |                                  | 0     |
| Observation Type | Continuous observation | 1 |
|               | Station Observation               | 2     |
| Date         | 0602                             | 0602  |
3. Result and conclusion
The integrated DB constructed through the precedent study must be created into the new table as the observed items revised. This is a very ineffective and difficult in DB management. Therefore, in this study, problems of the ecological/environmental integrated DB, which was constructed through the precedent study, were analyzed. And improvement was made to the ecological/environmental DB in order to conduct more effective management. The improved integrated DB is being applied as the base data to support the decision making for estuary management. As a result of applying the integrated DB, there was a significant increase in the efficiency in the DB management. First, it is not necessary to create new tables even if modified observed data enter as observation items. This is also useful when applying the integrated DB in the system. In case of numerous tables existing, the SQL becomes complicated within the estuary management system. On the other hand, in case of few tables existing, it is convenient since the only condition in the query have to be modified. However, when the tables become combined into a single table, the data searching speed will be decreased. The indices were created in order to solve the problems. The method utilized in this study can be applied to diverse fields. It can be applied in fields that require complex data structure related to atmosphere, ocean, and ecosystem.

Acknowledgement
This work was supported by the Ministry of Land, Transport, and Marine Affairs, under the research project entitled "Development of Integrated Estuarine Management System (20100051)" and under the supporting project to educate GIS experts.

Reference

[1] Seo J 2011 *Proc. Geospatial world forum*