Development of database design technique as part of the system of ecological monitoring of urban infrastructure objects

Larissa Gennadievna Gagarina, Aleksandra Igorevna Kononova, Yana Olegovna Teplova, Alexey Vladslavovich Gorodilov, Aleksander Mikhailovich Bain, Oleg Ivanovich Lisov, Victor Georgievich Dorogov, Catherine Georgievna Dorogova

National Research University of Electronic Technology, 4806 passage, 5 , Zelenograd, Moscow, 124498, Russian Federation
E-mail: yana.teplova@gmail.com

Abstract. The paper dwells on the problem of designing efficient database logical structures to improve the efficiency of processing data queries for solving the tasks of ecological monitoring and industrial ecological control. Computer network of ecological monitoring system with databases is represented as graph. The database design technique is suggested, which respects time characteristics of work as part of the system of ecological monitoring of urban infrastructure objects, based on graph theory. Implementation of logical database structures, designed taking into account time characteristics of the work, will allow avoiding unreasonably high time costs in information processing and analysis.
[Gagarina L.G., Kononova A.I., Teplova Y.O., Gorodilov A.V., Bain A.M., Lisov O.I., Dorogov V. G., Dorogova C.G. **Development of database design technique as part of the system of ecological monitoring of urban infrastructure objects.** *Life Sci J* 2014;11(11):571-573] (ISSN:1097-8135). http://www.lifesciencesite.com. 101

Keywords: ecological monitoring of urban infrastructure objects, ecological monitoring system, monitoring information processing, databases, graph theory

Introduction

Population growth, the concentration of resource flows in cities, large industrial centers and agglomerations inevitably lead to increased pressure on ecosystems and increase the risk of significant harm: environment and air pollution, deterioration of the living environment of people, etc. In order to reduce the technical and environmental, ecological and economic risks, as well as to forecast and make informed decisions in risk management ecological monitoring is needed, which determines levels of the environment contamination and maximum allowable concentrations of pollutants, etc. [1, 2, 3, 4]. There are several papers of J. Anderson, R. Gardner, W. Hommen, G. Suter, S. Bartell, J. Chang, concerning this field of research, as well as Russian scientists: V.A. Osipov, Yu.A. Komissarov, O.A. Drozdov, A.P. Shepelevsky, E.Yu. Bezuglaya, A.M. Gorshenov, etc.

Methods

Achieving these goals is not possible without the use of local, regional and other ecological monitoring systems (EMS), which solve the problems associated with the collection, processing and distribution of large volumes of information in real time using modern technology of geographic information systems, database management systems (DBMS) etc. The simulation of EMS workflow is used to solve the problems of managerial decision-making [5, 6, 7], then the efficient tools and techniques for the collection, compilation, storage and presentation of data are developed [8, 9].

There are several ways of solving the problem of dataware development for information systems in the area of ecological monitoring. One of the most commonly used is applying data warehouse technology. Datawarehouse is an object-oriented database specifically designed and intended for support decision making. Their use in ecological monitoring of is appropriate for the following reasons:

1) data consistensy;
2) single set of processes and business rules;
3) general semantics;
4) centralized and fully controlled environment;
5) datawarehouses can be easily created and filled up with data marts;
6) single metadata repository.

Main part

In particular, to improve the usage of data warehouses for the needs of industrial and ecological control at the production and ecological monitoring of urban infrastructure database design technique is proposed, based on graph theory. Dataware of a territorial EMS provides for a single computer network and the availability of multiple databases (Fig. 1) [4]. The main indicator of the efficiency of the monitoring information processing – total execution time of database load – is given by:

$$T = \sum_{P=1}^{P0} T_p^z + \sum_{S=1}^{S0} T_s^k \,, \quad (1)$$

where T_p^z – time of implementation of given set of queries, T_s^k – time of given set of update tasks, $P0$ – number of queries, $S0$ – number of update tasks.

Time of implementation the p-th query:

$$T_p^{\,z} = t_p^{pz} + t_p^{l} + t_p^{c} + t_p^{u} \,, (2)$$

where t_p^{pz} – running time to split the task (the query to view, add, edit, or delete records) to subtasks and control their implementation; t_p^{l} – running time of query to local database; t_p^{c} – running time to collect the arrays of intermediate data and generate a query result; t_p^{u} – time to initiate and transfer the query to local database as part of EMS information support.

Time of search required types of records:

$$t_p^c = (t_1^{acc} + t_1^{w}) + t_1^{ch} k_p^b \,, (3)$$

where t_1^{acc} – time of search required records in local database; t_1^{w} – waiting time to access and read data from database; t_1^{ch} – time to exchange between external memory and random access memory; k_p^b – number of records (blocks) to be read for p-th query.

Set of record types, placed on r-th node of EMS computer network and connected with a set of logical relations, is called a logical structure graph of r-th node. To determine the number of local databases in EMS computer network it is required to solve the tasks of normalization of the logical structure graph of r-th node, finding its disconnected and weakly connected components [10, 11].

Let $H^{(1)}{}_r = \{h^1_{pr}\}$ be a set of first-level record types of the logical structure graph of r-th node of graph G_r^{rec}, i.e. a minimal set of vertices of graph G_r^{rec} from which all vertices are reachable, or the base of the graph. Let $F(H^{(1)}{}_r)$ be a reachability set of base $H^{(1)}{}_r$. Then $F(H^{(1)}{}_r) = H_r$ and $\forall H_r^* \subset H_r F(H^*{}_r \neq H_r)$.

The reachability set for each record type are sets $H^{(1)}{}_r = \{h^1_{pr}\}$:

$$F(h^1_{pr}) = \{h^{(1)}_{pr}\} \cup Y^1(h^{(1)}_{pr}) \cup Y^2(h^{(1)}_{pr}) \cup ... \cup Y^n(h^{(1)}_{pr}) \quad (4)$$

Analyzing pairwise intersections of reachability sets, we find disconnected components of the logical structure graph of r-th node. Subgraphs G_{p1} and G_{p2} of the G_r^{rec} graph ($h^{(1)}_{p1r} \in H^{(1)}_r$ and $h^{(1)}_{p2r} \in H^{(1)}_r$, $h^{(1)}_{p1r}$, $h^{(1)}_{p2r}$ – the root vertices of the graphs) are disconnected if $F(h^1_{pr}) \cap F(h^{(1)}_{p2r}) = 0$.

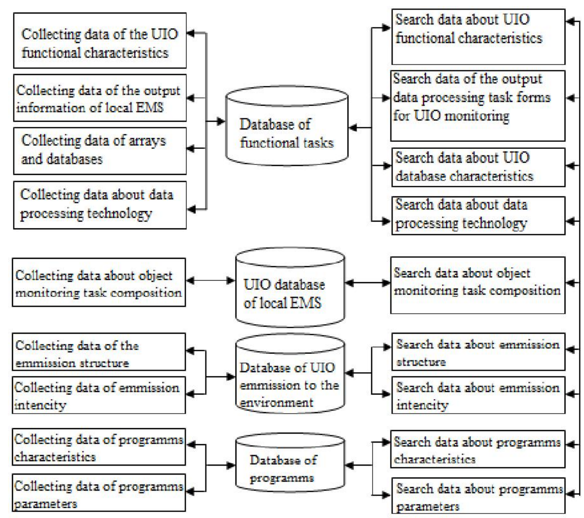

Fig. 1. Structural data schema of EMS of urban infrastructure objects

The set of record types and relations forming disconnected subgraph of the logical structure graph of r-th node form a subset, which is used in the design of the logical structures of local databases as part of the territorial EMS. The need to allocate weakly connected components of the logical structure of the r-th node justified by the requirements of the database efficiency and limitations of the amount of stored information.

The following technique of designing the database for EMS of urban infrastructure objects is suggested within the problem being solved.

1. Determining the set of record types of first level in hierarchy $H^{(1)}_{ri} = \{h^{(1)}_{pri}\}$ for i-th connected component of graph G_r^{rec} .

2. Determining the reachability set for each record type $H^{(1)}_{pri} \in \{h^{(1)}_{ri}\}$:

$$F(h^1_{pri}) = \{h^{(1)}_{pri}\} \cup Y^1(h^{(1)}_{pri}) \cup Y^2(h^{(1)}_{pri}) \cup ... \cup Y^n(h^{(1)}_{pri}) \quad (5)$$

3. Definition $M^{(p)} = \{m_j^{(p)}\}$ – sets of pairwise intersections by reachability set $m_j^{(p)} = F(h^1_{p1ri}) \cap F\{h^{(1)}_{p2ri}\}$.

4. Analysis of set $M^{(p)}$. If the set of pairwise intersections consists of empty subsets go to step 7, otherwise – to step 5.

5. Selection the minimum subset $m^{\min}$ of set $M^{(p)}$, $card(m^{\min}) = \min[card(m_j^{(p)})]$, whose elements compose a set of vertices of selected subgraph of graph G_r^{rec}.

6. Eliminating vertices of selected subgraph and lines leading to its root node from the set of vertices of graph G_r^{rec}; proceed to step 2.

7. Adding disconnected components of the graph G_r^{rec} obtained by eliminating vertices and lines to the set of allocated subgraphs.

Conclusion

Implementation of logical database structures, designed taking into account time characteristics of the work, will allow avoiding unreasonably high time costs in information processing and analysis.

The described database design technique as part of a territorial EMS has been tested during an ecological monitoring of municipal facilities in Zelenograd administrative district of Moscow. Reliable results have been obtained.

Corresponding Author:
Dr.Gagarina Larissa Gennadievna
National Research University of Electronic Technology
4806 passage, 5 , Zelenograd, Moscow, 124498, Russian Federation
E-mail: yana.teplova@gmail.com

References

1. Gagarina, L.G., P.Yu. Chumachenko and A.M. Bain, 2008. Some features of the design of decision support systems in a specific domain. Defense industry achievements – Russian scientific and technical progress, 1: 24-26.
2. Chan, K.Y. and L. Jian, 2013. Identification of significant factors for air pollution levels using a neural network based knowledge discovery system. Neuricomputing. Neuricomputing, 99: 564-569. DOI: 10.1016/j.neucom.2012.06.003
3. Gacquer, D., V. Delcroix, F. Delmotte and S. Piechowiak, 2011. Comparative study of supervised classification algorithms for the detection of atmospheric pollution. Engineering Applications of Artificial Intelligence, 24: 1070-1083. DOI: 10.1016/j.engappai.2011.04.002
4. Teplova, Ya.O., 2011. Development of a theoretical framework for creating the software in the field of ecological monitoring of production facilities. Research in the field of data processing, storage, transmission and protection: paper key points. National Conference with international participation, 25-28 October 2011, Moscow, Exhibition of Achievements of National Economy, pp: 167.
5. De Mesnard, L., 2013. Pollution models and inverse distance weighting: Some critical remarks.. Computers & Geosciences, 52: 459-469. DOI: 10.1016/j.cageo.2012.11.002
6. Ma, Y.J., Y.K. Guo, D. Silva, O. Tsinalis and C. Wu, 2013. Elastic Information Management for Air Pollution Monitoring in Large-Scale M2M Sensor Networks. International Journal of Distributed Sensor Networks, 2013. Date Views 9.06.2014 dx.doi.org/10.1155/2013/251374. DOI: 10.1155/2013/251374
7. Singh, V., C. Carnevale, G. Finzi, E. Pisoni and M. Volta, 2011. A cokriging based approach to reconstruct air pollution maps, processing measurement station concentrations and deterministic model simulations. Environmental Modelling & Software, 26: 778-786. DOI: 10.1016/j.envsoft.2010.11.014
8. Aljumaily, H., D. Cuadra and P. Martinez, 2014. Applying black-box testing to UML/OCL database models. Software Quality Journal, 22: 143-184.
9. Wang, H., L. Li and Yu. Jiao, 2014. A relationship-based and object-oriented software for monitoring management during geotechnical excavation. Advances in Engineering Software, 71: 34-35.
10. Portnov, E.M., 2009. Technique of determining the actual recording time of discrete events in information and control systems. Defense industry achievements – Russian scientific and technical progress, 4: 27-32.
11. Dubovoy, N.D., V.V. Slyusar and Ya.O. Teplova, 2012. Technique of determining the actual performance of information and control systems in power engineering. Defense industry achievements – Russian scientific and technical progress, 3: 8-12.

7/10/2014