Features of the implementation of olap-systems in environmental monitoring of the marine environment

G.D.Kaziyeva[1], A.E. Abzhanova[2], S.K. Sagnayeva[1], G.K.Sembina[3], T.K.Ermagambetov[4]

[1]Sh. Yessenov Caspian State University of Technology and Engineering, Aktau, Kazakhstan

[2] L.N. Gumilyov Eurasian National University, Nus-sultan, Kazakhstan

[3]International University of Information Technology, Almaty, Kazakhstan

[4]Aktobe Regional State University named after K.Zhubanov

Abstract. For complex systems, it is necessary to constantly evaluate the influence of control actions on the conditions of the system, to be able to record the changes in the conditions of the system and evaluate the degree of influence of various factors on the conditions of the system. With the complexity of the data structure and its volume, there is the question of information processing arises, where solution is impossible without the involvement of information systems and technologies. One of the problems is maintaining an acceptable system response time to the request. In multidimensional data analysis technology OLAP (Online Analytical Processing) slices or representations of multidimensional cubes are used as answers to user requests to solve the problem.

The article discusses the modeling technology of complex TOFI systems (object type, relations between objects, factor, measure) developed at the Factor System Research Company (Astana), as an OLAP system for monitoring the quality of the Caspian sea waters. The TOFI technology was used to develop the Caspian environmental monitoring information system. Our task is to adequately describe the state of this system using its properties and indicators, taking into account the influence of all necessary factors on the state of the system.

Keywords: OLAP, TOFI, information model, multivariate analysis.

1 Data Concept

1.1 TOFI Technology

The TOFI technology was developed by the Kazakhstan company KSI Factor (Nur-Sultan), the technology includes a methodology for researching the subject area and at the same time offers tools to automate the proposed methodology.The name of the technology is formed from the first letters of the main entities of this technology: "Type of object", "Relationship between Types", "Factor", "Meter". The basis of the TOFI is its information storage, which actually consists of two levels - relational and multidimensional. At the relational level, the physical structure of data storage is described. At a multidimensional level, a logical multidimensional representation of data for system users. [1].

This approach allows you to determine the methods and steps subject area studies, identify the main entities for building a subject area model. The technology is intended to build a model of the subject area as a complex organizational system from the point of view chosen by the researcher, as well as to study and identify new knowledge about the subject area, in accordance with the purpose of the study. The subject area model built using the TOFI technology is called the TOFI model.

ICID-2019 Conference

Data is the main object that both users and TOFI applications are work with. Data in TOFI technology refers to a set of elements (O, P (s, q), T), where O (Owner) - data owner, P (Feature) - feature, s (Status) - data status, q (Author, Provider) - data provider, T (Time) - period or current data interval [2].

Objects and relationships between objects are data owners. The relationship between objects is a set of objects that has its own features and has the behavior inherent in objects. The concept of relations between objects (which are inherited from relations between types of objects) turned out to be a rather flexible tool in modeling a subject area.

Features are a separate TOFI object, independent of objects and relationships between objects. Features are created from feature prototypes. Prototypes of features are measurements, attributes, factors, objects, relations between objects, units, periods.All prototypes are separate objects. Measurements model the quantitative features of objects and represent a multidimensional cube, the dimensions of which are factors. Attributes are features, whose values are the qualitative features of objects: strings, date, time, file, multi-line text, masks. Factors, objects, relationships between objects, units of measure - these are qualitative features of objects. This structure reflects the TOFI data concept.

This set defines the only features value of P for owner O in the time interval T (the only features value can be a list or hierarchy if the feature is an ambiguous quality feature or a complex feature).

Thus, the basic concepts of technology are the following:

- Object - any object of the “mini-world” or the essence of the subject area under consideration. Objects can be grouped into “object types”. An object in TOFI technology can be either the owner of the data or the feature itself, when one object is a feature of another object. For example, the «Meter Group» feature may be the «Object Types» feature.

- Object type - a selected set of features that describes the state of specific objects of this type. The Object type feature represents a group for specific objects and determines all their qualitative and quantitative features.

- The relations between the types of objects is a collection of types of objects that are related to each other in a certain way. Relations between types of objects are also a kind of "type of object." Sometimes a particular feature is not a features of one object, but more than one object at a time. In such cases, it is convenient to project these several objects into the same relationship between the objects.

- Relations between objects are similar to relations in relational databases and represent a set of objects and relations between objects. An instance of a relationship of an object type consists of the participants of the object that are part of the relationship. The relationship between the objects acts as the owner of the data and as a feature in the TOFI technology, in case if the relationship between the objects is a feature of another object or the relationship between the objects.

- A feature is a certain characteristic of an object or a relationship between objects that determines its state. Features are divided into quantitative (“Measurements”), qualitative (“Factor”) and other (“Other features”).

- Units are a measure by which quantitative data are measured. Units may, in turn, be feature. This is possible when the counter changes its unit of measure from one object to another. In this case, a measurement-based feature inherits its unit of measurement from an object or from a relationship between objects, and the corresponding unit of measurement is another feature of the object or relationship between objects.

- A factor is a qualitative feature of an object or a relationship between objects of a subject area, which is a list of possible states. This feature has no units.
- A measurement is a quantitative feature of an object or a relationship between objects in a subject area that has a unit of measurement.

Attributes are all the feature of an object or the relationship between objects that are not factors or meters. These include graphical objects, diagrams, descriptions of some process plans, the location (address) of objects, and others.

2 Designing an information system

2.1 TOFI Information Model

The designing of information systems for environmental monitoring is associated with the features of environmental management. For example, while designing a system, it is necessary to take into account the nature of data, the dynamics of biological processes, and, on the other hand, take into account the influence of industrial society on the state of resources [3,4]. Some experts argue that to monitor the state of the environment, with taking into account its interaction, significant information resources are needed. Such an approach assumes the existence of a special methodology based on a specific model, probabilistic approaches to management and specific information systems of a statistical and geoinformation nature.

Traditionally, environmental observations of the Caspian Sea are based on indicators such as: hydrometeorological observations (wind strength; air temperature; water temperature); hydrological observations (determination of salinity, temperature, turbulence and transparency of water; determination of oxygen mode (dissolved oxygen content); measurement of hydrogen (pH); measurement of water depth), hydrochemical observations (biogenic substances (ammonium nitrogen), nitrate nitrogen, nitrite nitrogen total nitrogen, total phosphorus); heavy metals (Cd, Cu, Pb, Zn)), hydrobiological observations (taxonomic structure, quantity and biomass, structure of dominant types, abundance and biomass of the main groups and types, distribution over the water area of the reservoir; basic indicators of the ecological state of the reservoir) [5,6].

Tools for building a TOFI model include entities for describing model features: qualitative features of domain objects are described using the «Attribute features», «Factor» entities, quantitative features are the «Measure» entity, additional features can be described by auxiliary entities («Units of Measurement»), («Periods», «Scales») and entities «Other properties». (see Fig. 1)

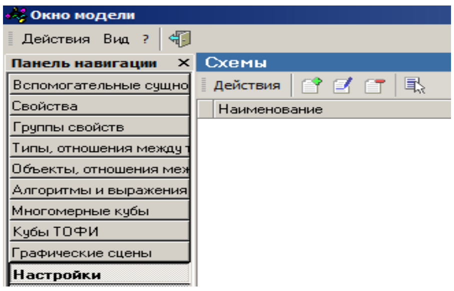


Fig. 1 TOFI model window

The TOFI toolkit allows you to add periods to the system: 2014, 2015, 2016. The "Attribute features" entity allows you to enter a description of such features as Latitude and Longitude. The “Factor” entity with the name “Square”, having the values shown in Figure 2, determines the type of the “Station” object.

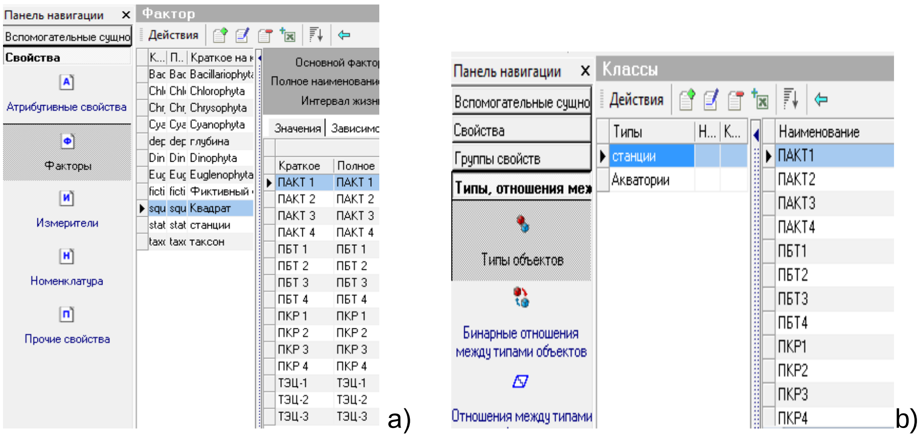


Fig. 2 Cluster factor «Square» (a) and object Type «Stations» (b)

The quantitative features of the model are described by the essence of the “Meter” and meters will be considered that describe the chemical and physical features of water such as total hardness, carbonates, bicarbonates, total alkalinity, suspended solids, etc. These meters are combined in a characteristic group called "General rigidity».

The technology of complex multivariate data analysis is implemented in TOFI using standard TOFI cubes. The «Multidimensional Cubes» tool uses to build a multidimensional cube. A cube prototype is pre-built and two cube measurements are selected. Two measurements “years” and “stations” are selected as measurements for a cube, with the measurement type “Periods” and “Objects and relations between objects”, respectively.

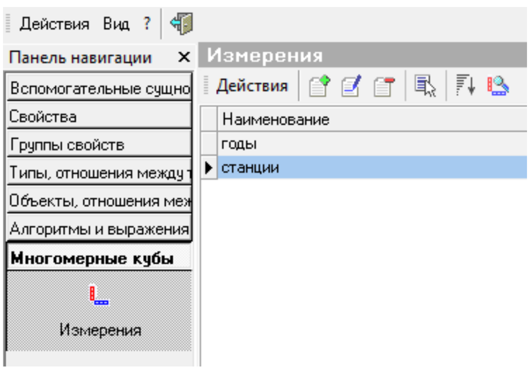


Fig. 3 Cube Measurements Creation Window

A procube can have any number of measurements (≥2) and is a hierarchical tree of groups of meters and indicators.

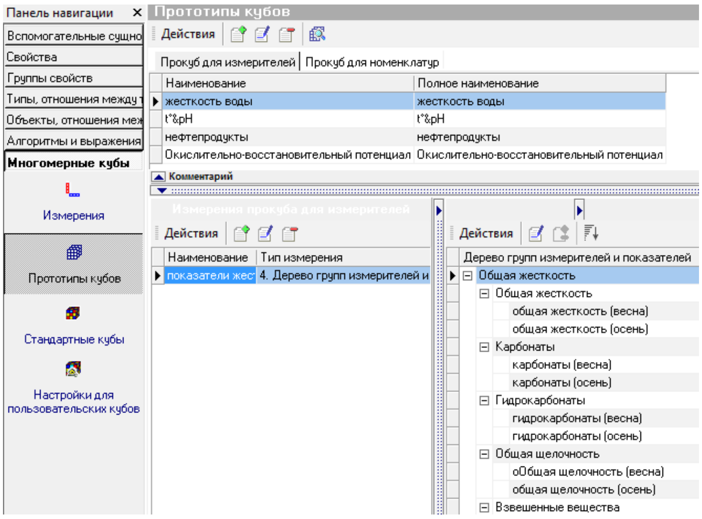


Fig. 4 Procube «Water hardness»

The procube view allows you to see a table with a data structure that is associated with the selected dimensions. A procube contains no data and the "Values" column is filled with zeros.

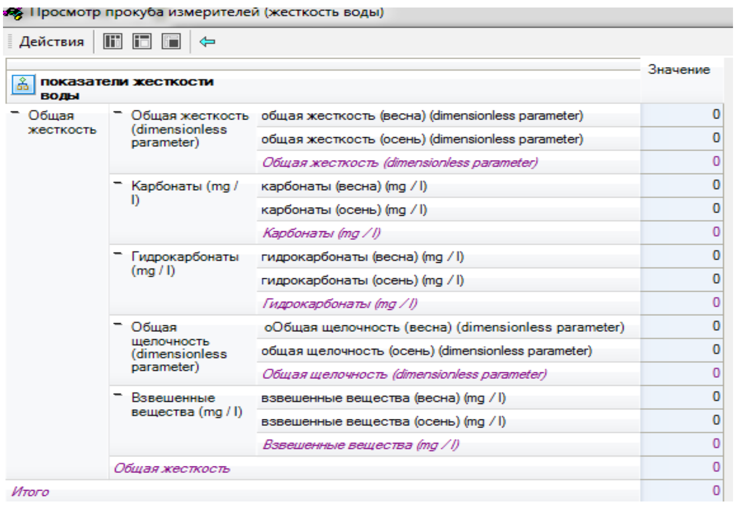
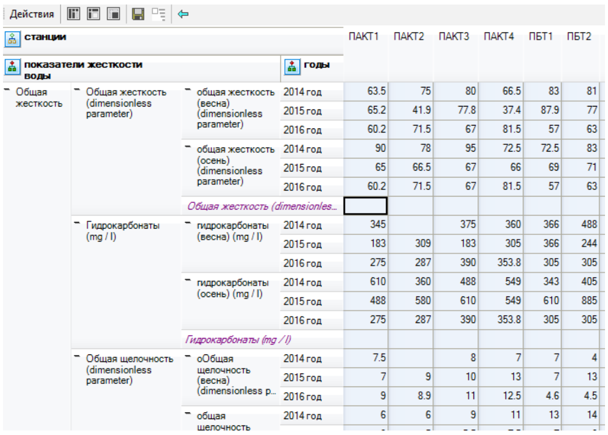


Fig.5 Viewing a procube of a standard cube

To enter data, the “Data Window Manager” tool is used, which allows you to enter data manually or through importing from Excel. The cube representation allows you to view data in a user-friendly way.



Действия

станции / показатели жесткости воды			годы	ПАКТ1	ПАКТ2	ПАКТ3	ПАКТ4	ПБТ1	ПБТ2
Общая жесткость	Общая жесткость (dimensionless parameter)	общая жесткость (весна) (dimensionless parameter)	2014 год	63.5	75	80	66.5	83	81
			2015 год	65.2	41.9	77.8	37.4	87.9	77
			2016 год	60.2	71.5	67	81.5	57	63
		общая жесткость (осень) (dimensionless parameter)	2014 год	90	78	95	72.5	72.5	83
			2015 год	65	66.5	67	66	69	71
			2016 год	60.2	71.5	67	81.5	57	63
		Общая жесткость (dimensionles...							
	Гидрокарбонаты (mg / l)	гидрокарбонаты (весна) (mg / l)	2014 год	345		375	360	366	488
			2015 год	183	309	183	305	366	244
			2016 год	275	287	390	353.8	305	305
		гидрокарбонаты (осень) (mg / l)	2014 год	610	360	488	549	343	405
			2015 год	488	580	610	549	610	885
			2016 год	275	287	390	353.8	305	305
		Гидрокарбонаты (mg / l)							
	Общая щелочность (dimensionless parameter)	оОбщая щелочность (весна) (dimensionless p...	2014 год	7.5		8	7	7	4
			2015 год	7	9	10	13	7	13
			2016 год	9	8.9	11	12.5	4.6	4.5
		общая щелочность	2014 год	6	6	9	11	13	14

Fig.6 Standard cube «Stations»

If necessary, you can change the structure of the cube, for example, in the original cube, you can change the arrangement of rows and columns in accordance with Figure 7.

Представление куба (Жесткость воды)

Действия

показатели жесткости воды		Общая жесткость					
		Общая жесткость (dimensionless parameter)			Карбонаты (mg / l)		
станции	годы	общая жесткость (весна) (dimension...	общая жесткость (осень) (dimensi...	*Общая жесткость (dimensionless p...*	карбонаты (весна) (mg /...	карбонаты (осень) (m...	*Карбонаты (mg / l)*
ПАКТ1	2014 год	63.5	90		0	0	
	2015 год	65.2	65		0	0	
	2016 год	60.2	60.2		0	0	
ПАКТ2	2014 год	75	78		0	0	
	2015 год	41.9	66.5		0	0	
	2016 год	71.5	71.5		0	0	
ПАКТ3	2014 год	80	95		0	0	
	2015 год	77.8	67		0	0	
	2016 год	67	67		0	0	

Fig.7. New cube view

3 Compliance of requirements to OLAP systems.

Consider the requirements of E. Codd to implementation tools for OLAP systems and their implementation in TOFI [7].

1. Multidimensional data representation *(Multi-Dimensional Conceptual View).* A logical multidimensional model is constructed in the form of multidimensional cubes, based on the relational model in TOFI.
2. Transparency. The user in the TOFI does not get a concept what a specific means are used to store and process the data, how the data is organized and where they comes from. This

is achieved by using the TOFI toolkit and the navigation bar. “Multidimensional cubes” (consisting of the “Measurements”, “Prototype cubes”, “Standard cubes” panels).

3. Accessibility. Measurement tools of prototype cubes and standard cubes allows to select and communicate with the best data source from the TOFI repository to a given request for forming a response.
4. Consistent Reporting Performance. The formation of a cube of any type in TOFI is independent of the number of its measurements, since the prototype cube is being pre-built, and after that are «adding» the measurements of the periods and measurement of the objects/relationships. An increasing of the number of measurements does not entail complication of the cube views for the analytic user.
5. Client-Server Architecture support. *(Client-Server Architecture).*TOFI technology is both a client and server OLAP tool. The cube itself and the storage of aggregate data is performed by the server, and the client application has the ability to perform part of operations on cube.
6. Equality of all measurements *(Generic Dimensionality).* This requirement is not fully met, since measuring of the periods and measuring of the objects/relationships are mandatory for TOFI cubes.
7. Dynamic Sparse Matrix Processing *(Dynamic Sparse Matrix Handling).* In TOFI technology, sparse matrices store undefined values in the form of empty cells, which allows to maintain access speed regardless of the location of the data cells for cubes which are having a different number of measurements and different sparseness of the data.
8. Support for multi-user data mode *(Multi-User Support).* The TOFI client-server architecture provides simultaneous access for a large number of users. Moreover, the analysis is carried out on different aspects of information, on different slices of cubes, regardless of the structure and measurements of the cubes. TOFI provides data integrity and protection.
9. Unrestricted Cross-dimensional Operations. All multidimensional TOFI operations are uniformly and consistently applied to any cubes with any number of measurements.
10. Intuitive Data Manipulation. Implementation tools for multi-dimensional TOFI cubes allows to swap rows and columns of a table, hide / expand the measurements hierarchy, etc. (fig. 3)
11. Flexible Reporting. TOFI tools support various ways of visualizing data in both tabular and graphical forms. A mechanism for saving reports is being developed in any user-friendly form.
12. Unlimited Dimensions and Aggregation Levels. E. Codd himself in the rules recommended a up to 15 cube measurements. Theoretically, in TOFI there are no restrictions on the number of supported dimensions, but practically cubes with more than 10 measurements has not been created yet.

4 Conclusion

Currently, a few analytical systems have been developed, designed using OLAP technology (Hyperion OLAP, Elite OLAP, Oracle Express and others). The choice of the OLAP system as a platform of TOFI allows to develop the structure of the information model and use the TOFI for data entry.

To obtain slices of information according to monitoring data was used the tool "multidimensional cube" TOFI. The cube representations are some selection based on data from the TOFI repository. In case of using the generally accepted relational approach to obtain such a sample by the hydrological and physicochemical parameters of the Caspian, it would be necessary to write a complex query to the database. In case of using the multidimensional TOFI cube, the data processing procedure becomes transparent and clear, which allows the user to accelerate decision-making in accordance with the purpose of monitoring.

References

1. Gabbasov, M.: Architecture and Capabilities of TOFI Technology. In: Journal "Computing Technologies," Part 1. Joint issue of the Journal of East Kazakhstan State Technical University named D. Serikbayev and the journal "Computing Technologies" of the Institute of Computational Technology Siberian Branch of the Russian Academy of Sciences- Ust-Kamenogorsk, 270 – 277(2013).
2. Gabbassov M.: TOFI technology capabilities for data processing and visualization. In:Conference collection Application of Information and Communication Technologies-AICT2014, 276 – 277, Astana (2014)
3. Karlykhanov O.K.,Bekbaev R.K., Tazhieva T.C.Water problems of Aral sea and ways of their solution. In: World Applied Sciences Journal, 31(6), 1214-1216 (2014).
4. Neverova G.P., Zhdanova O.L., Kolbina E.A., Abakumov A.I. A plankton community:a zooplankton effect in phytoplankton dynamics. In: Computer Research and Modeling, 11(4), 751-768(2019).
5. Agnew, D.:In: Review: the CCAMLR Ecosystem Monitoring Program. Ant. Sei, 9 (3):235-242 (1997).
6. Ismailova, A.A., Zhamangara, A.K., Kaziyeva, G.D., Abakumov, A.I., Park, S.Y., Sagnayeva S.K. Technologies of information monitoring biogens lakes of Kazakhstan. In: News of the National Academy of Sciences of the Republic of Kazakhstan, Series of Geology and Technical Sciences, 3 (430), 69-73, (2018).
7. Zhuat M.B.. Bakayev D.D.. Sagnayeva S.K. Olap v tekhnologii TOFI// Primeneniye matematicheskogo modelirovaniya i informatsionnykh tekhnologiy v issledovaniyakh sotsialno-ekonomicheskikh problem: Materialy nauchno-prakticheskoy konferentsii. posvyashchennoy 50-letiyu k.f-m.n. dotsenta Gabbasova M.B.- Astana. NII EITT, 172-177 (2011).