Glava M. G., Dr. Sci. Malakhov E. V.

Odessa National Polytechnic University, Odessa I. I. Mechnikov National University, Ukraine

THE INTEGRATION OF INFORMATION SYSTEMS ON THE BASIS OF INFORMATION MODELS UNION OF THE SUBJECT DOMAINS

Глава М. Г., д.т.н. Малахов Е. В.

ИНТЕГРАЦИЯ ИНФОРМАЦИОННЫХ СИСТЕМ НА ОСНОВЕ ОБЪЕДИНЕНИЯ ИНФОРМАЦИОННЫХ МОДЕЛЕЙ ПРЕДМЕТНЫХ ОБЛАСТЕЙ

Owing to the advances in the informatization of modern society, most organizations when they begin business process reengineering already have some ISs, which eventually also demand reengineering. Due to such reasons there are arising a difficulties of the connection and analyzing the contained information in the miscellaneous sources. It is becoming urgent task of consolidating the existing information systems. Moreover, in this case, the integration of information systems (IS) is actual not so much on the external schema level, i.e. on the user view, as at the data level, actually the part of the IS, which is the information model of the subject domain.

Analyses of existing data integration approaches outcomes in the fact that in order to minimize the data duplication (in terms of storage) and time (to receive information by the user), it is necessary to identify common information elements of all databases. This also require mathematical integration of the IS domain (ISD) models, describing each of the combined databases.

There are several alternative layers of data integration: integration at the physical (converting data into the uniform format), syntactically or logically (based on the resemblance of merged data; considers structural properties of data from various sources) and semantic (based on the similarity of the merging data) layer.

In the context of our research, the most promising level of the integration is the semantic layer, which allows to integrate not only structurally identical data, but also having the same meaning. Let us note here, that universal approach to IS integration at semantic layer of the data is under development now.

After analyses of existing solutions for the ISs integration, it can be concluded that considered methods do not complete solve the problem of data integration. The problem of databases integration on the conceptual layer remains actual, this why this paper proposes an approach for the integration of relational databases as domain models. The approach will allow us to automate the creation of conceptual schemas for multiple layers of ISD, which will give an opportunity to save work and time costs.

The procedure for combining domain models depends on the similarity of information systems domains. For the ISDs, which describe different business domains, an integration is possible only for the typical objects (defining objects) representatives of many ISDs (e.g., 'contractors', 'employees' etc.). For ISDs, reflecting the same kind of business, we need find matching entities and integrate them. For domains, that describe the same area of business should be matched both the objects that nondefining the domain and objects which can belong to both compared domains, so-called "borderline" objects.

The author [2] presented the technology to find the projections of the same universal entities on ISD, where objects are compared on the base of

the values of properties of the objects' instances. Algorithms for comparing depends on the data type of specific properties.

Accordingly to proposed technology, objects of the potentially similar ISD must be prepared for the matching: allocate the essential properties, based on the information amount of each property and expert assessment; rank the objects of each compared ISD by significance, based on the number and importance of a particular object relationships with others in the same ISD and the number of significant properties, measured by a certain scale (ordinal, nominal, numeric); sort corteges according to the values of ordinal and nominal properties, keeping obtained previously properties rank.

Then the ratio the of objects' similar properties to their total number can be determined. The resulting ratio is compared to a threshold, given by an expert and concludes, whether the compared objects are the similar.

Finally, a common model of information systems domains is proposed by including from the most similar objects the most similar properties and expanding them by different properties, ordered by importance.

Proposed approach allows to automate the development of the ISD model (conceptual schema), reducing the work effort and time required for the database integration. It also allows reducing the number of problems to be solved by experts.

REFERENCES

1. Васильева Т.П., Глава М.Г. Основные проблемы и методы интеграции баз данных // Первый независимый научный вестник.– 2015.– № 1.– C. 28–32.

2. Glava M. Searching Similar Entities in Models of Various Subject Domains Based on the Analysis of Their Tuples / M. Glava, E. Malakhov // 2016 International Conference on Electronics and Information Technology (EIT).– 2016.– P. 97-100. doi: 10.1109/ICEAIT.2016.7501001.