

Natalya SHAKHOVSKA¹, Mykola MEDYKOVSKY¹, Liliana BYCZKOWSKA-LIPIŃSKA²

Lviv Polytechnic National University (1), Wyższa szkoła informatyki w Łodzi (2)

The main areas of dataspace realization

Abstract. Problems which arise up during work with separate information sources and databases are analyzed.

Streszczenie. Problemy, które pojawiają się podczas pracy z oddzielnymi źródłami informacji i baz danych są analizowane.

Keywords: Dataspace, Datawarehouse, Integration Methods, Consolidation, Information Product.

Słowa kluczowe: Dataspace, Datawarehouse, integracji metody, Konsolidacja, Informacje o produkcji.

The physical system is dynamic and its elements have evolved at different rates. It complicates the collection and processing of information on elements of such a system. To work with different types of information from different sources, we can apply dataspace.

Traditionally, experts used usual for them sources of information for solving tasks. Apparently, this approach has incomplete information, which is processed. Many sources of data and services that exist on the Internet are causing the need for a radical change in the methods of getting data. This change is a tasks, which is formulated independently of existing data sources. After its formulation the identification of relevant sources, bringing data to the appropriate type, integration, identification services which allow solving a separate part of tasks should be carried out. The adoption of adequate solution require the data, coming from different sources to satisfy the following requirements: be complete, consistent and received on time; be informative, because they should be applied for decision support; be of uniform structure for the opportunity of being downloaded in single datawarehouse and analyzed; kept in uniform models of data and be independent of the development platform for the opportunity of using this data in other means. But today there are no data processing methods that would satisfy all of the requirements for data processing.

Datawarehouse is an aggregated information resource that has consolidated information from all areas of concern and is used for decision support.

Dataspace is a set of all information product domain

$DS = \langle DB, DW, Wb, Nd, Gr \rangle$,

where **DB, DW, Wb, Nd, Gr** are information products that submit a set of databases, datawarehouse, web pages, text files, spreadsheets, image data respectively. In an energy system databases, datawarehouses, text files, spreadsheets that are described in different formats are used.

Talking about an information product, we mean its content (information resource, IR), and the set of information about it (accommodation, access scheme, speed of information update, etc.). We are also interested in the operations which are carried out over IR depending on its DSIR. The main task of dataspace is to allow the user to work with data sources without knowing its DSIR, accommodation, access methods, etc.

Consolidated data is derived from multiple sources and systematically integrated heterogeneous information resources, which together are have such features as completeness, integrity, consistency and adequacy/ This consolidated information is model of the subject area for its analysis and processing efficiency in the processes of decision making.

The dataspace requires much more technological and methodological solutions, as it contains processed information from various data structures, uncertain advance, and use different tools for processing and storage.

If we look at technologies that will help realize the potential of data space, we first need to focus on grid and cloud computing. Cloud architecture is opened for access

through the network, not only within the grid. Reference to pools with computing resources and storage data systems are made by the standard protocols, such as WSDL and SOAP, or using technology Web 2.0 (REST, RSS, AJAX), as well as through the existing technology grid.

Dataspace is typically used for decision making in a particular industry based on primary data. Examples of such areas are:

- an educational process (availability of operationally-independent and dependent structural information products for accounting students, analyzing their performance, recording and analysis of scientific publications for students and teachers, the formation of scientific schools, etc.);
- tourist area (the existence of different, unrelated between information products such as tour operators, information agencies, management of public tourism sector);
- medicine (the flows of information coming from different sources with unknown structures, the relationship between data that can be used by the processes of decision support, is not fully known);
- personal information management (management of personal computers without having to download a program and view the contents of information products with possible extensions to mobile devices);
- web-community manage (the analysis of user profiles in social networks and their several similar configuration to improve the perception of information by the user);
- monitoring and processing of news (there are usually adapted semi-structured text files, there is a problem in determining likelihood of events based on analysis of sources that these events described);
- annotation and summarization system (there are used semi-structured text files, are built document semantic network, whose peaks are compared with certain objects of the database to determine the weight of the term and sentence, in which he is included in the final essay);
- analysis of scientific sources (searching in databases that contains information about the quality of faculties, in scientific articles or research, comprehensive analysis of the listed earlier sources will permit to solve such problems as defining promising scientific approach, unique works of scientist, use data for interdisciplinary studies);
- energetic (electricity plants use different software, data about some of the sensors may come later, searching for information on grouped data with power stations, accounting of these stations is relevant).

REFERENCES

- [1] Donald Kossmann, Jens-Peter Dittrich. Personal Data Spaces. (2007) http://www.inf.ethz.ch/news/focus/res_focus/feb_2006/index_DE.

The correspondence address is:
e-mail: natalya233@gmail.com