

An Integrated Environment of Handling and Storing Information in the Tasks of Inductive Modeling

Nataliya V. Shcherbakova, Volodymyr S. Stepashko

*International Research and Training Centre of Information Technologies and Systems of the NAS and MES of Ukraine,
Glushkov ave., 40, Kyiv, 03680, Ukraine*

nataliya.shcherbakova@gmail.com, stepashko@irtc.org.ua

Abstract. *When solving real tasks of model construction from statistical data, the question arises of storage of and providing effective access to the information. On the stage of input data processing there are typical difficulties which do not depend on the choice of a modeling method, namely data processing in different formats which contain omissions and untypical small values etc. From the other side, the question of output information storage exists like structure and parameters of models, estimation of validity and exactness, graphics etc. To solve such kind of problems, the integrated environment of information storage is developed which allow to structure input data of different types and to use the information already existing in the base and also provide the storage of a complete information on experiments and results of calculations. To operate with the algorithms used for inductive modeling, the use of the XML-storage for saving of statistical input data and calculations results is relevant. Keeping metadata in a relational database is appropriate in this case, which allows simplifying any manipulations with them.*

Keywords

Inductive modeling, integrated environment of storage, information structuring, XML-storage.

1 Introduction

Algorithms of inductive modeling are actual for solving of practical tasks in the area of modeling of economical, ecological, and other processes [1, 2]. A necessity in accessible storage and drawing on scientific researches ripened already a long ago.

To solve the existing problems, an integrated environment of information storage would help allowing to structure input data of different types and sources, use input data already existing in a base, and also providing storage of complete information on experiments and results of calculations.

The main task is generalization and structuring of basic formats of input information, developing the standardized storage formats for calculation results and designing on this basis an integrated environment of information storage for operation with inductive modeling algorithms.

The offered model of data storage is intended for working out problems of storage of input statistical data and results. Developed architecture of the system of information storage in the tasks of inductive modeling gives a possibility to develop a software system that will allow freely manipulating with existent information and adding new one.

2 Functional requirements to system

The system must allow to import and export an input data, storage and processing an input data, storage an output data with complete information about calculations, to form reports by results of calculations. Let's examine more detailed each of functional requirements to system.

In the first place, the system must execute the import of data from the different types of sources like files in a text kind, xls-tables, and other. One of variants of expansion of the system can be extension the spectrum of the used sources, for example import data from other databases or popular information from the network of Internet.

On the stage of importing data must be executed the primary processing of data, namely addition of omissions, replacement of anomalous, atypically large or small values et al operations that prepare data for the use in algorithms.

Subsequent expansion of the system, which touches the primary processing of data consists at first in extension the types of sources and also in passing of statistical data to other class of bases functions.

Calculations results must be storing in the environment in the standardized kind which will allow to build are strictly formalized reports on results.

Important is storage of complete information about calculations, including information about an user, date and time of implementation of actions, which one actions were executed, with pointing of the used files and place of their finding.

The function of export must allow executing transformation of format of data storage from a environment in necessary for the use in already existent realization of algorithms of inductive modeling. That is to execute back to the export procedures, namely storing data is in a text format, or to any other. So exactly as well as module of import of expansion of the system in extension the spectrum of the used formats of storing information.

It should be noted that the system must work as with algorithms executed in an environment so for outside him. Algorithms are executed in an environment are such algorithms which work with data in those formats in which they are stored in a environment. Algorithms which are executed outside environment require transformations of information for the subsequent use.

Last requirement which is enough important for such level systems consists possibility of subsequent expansion in due to extension functionality of the system, or perfection existing functions.

3 Architecture of system

For work with the algorithms of inductive modeling, optimum will be use a XML-storage for storage of statistical input data and results of calculations. XML data can be storing in a two-dimensional file, object-oriented or to the relational database. In the case of work with information in the algorithms of inductive modeling, storage of data in a two-dimensional file is more comfortable than traditional SDBM [3]. It should be noted that the use of XML-storage allows to give information for the use in the modules of realization of algorithms, especially when it is necessary to process data of applied area with difficult hierarchical copulas.

One of the most ponderable requirements to the system that is possibility of storage of complete information about the calculations, performer, date and time, what algorithms were used etc. Such information can be named metadata. Her appropriately to keep in relation SDBM, that allows due to their typical mechanisms which acquired wide spreading, to carry out fast access to information etc [4, 5]. A database contains complete information about the executed operation, performer, time, files which was involved, and place of their physical location, formats of files, short comments by which it is possible to work with a document, and main attributes of document for which the search of document is carried out in a XML-storage.

For granting a possibility to work with data of environment to end users, optimum to use a web-interface, that considerably simplifies work with the algorithms integrated in an environment, to receive and storage of results and reports in a kind necessary for the user. The use of other subsystems of environment is thus simplified also, namely to import data to the environment and export data for the use in already existent realization of algorithms of inductive modeling and etc.

The system must consist of the following key subsystems:

- subsystem of primary handling and data import, which allows to standardize input data, which are imported from different sources, namely *.xls, *.txt, *.csv files in a XML-storage, and to conduct the primary processing of data, which will allow to fill omissions and liquidate atypically small or large values etc.;
- subsystem of data handling and use of algorithms, which is main part of environment, and keeps realization of algorithms and enables their use;
- subsystem of reports forming and storage of results of calculations, allows to give results in a necessary form, namely parameters of model, estimation and other, and also to place them in a XML-storage;
- a subsystem of export data is for the use in the existent products of software's, that allows to export data from an environment in a necessary format for the use in the already realized algorithms.

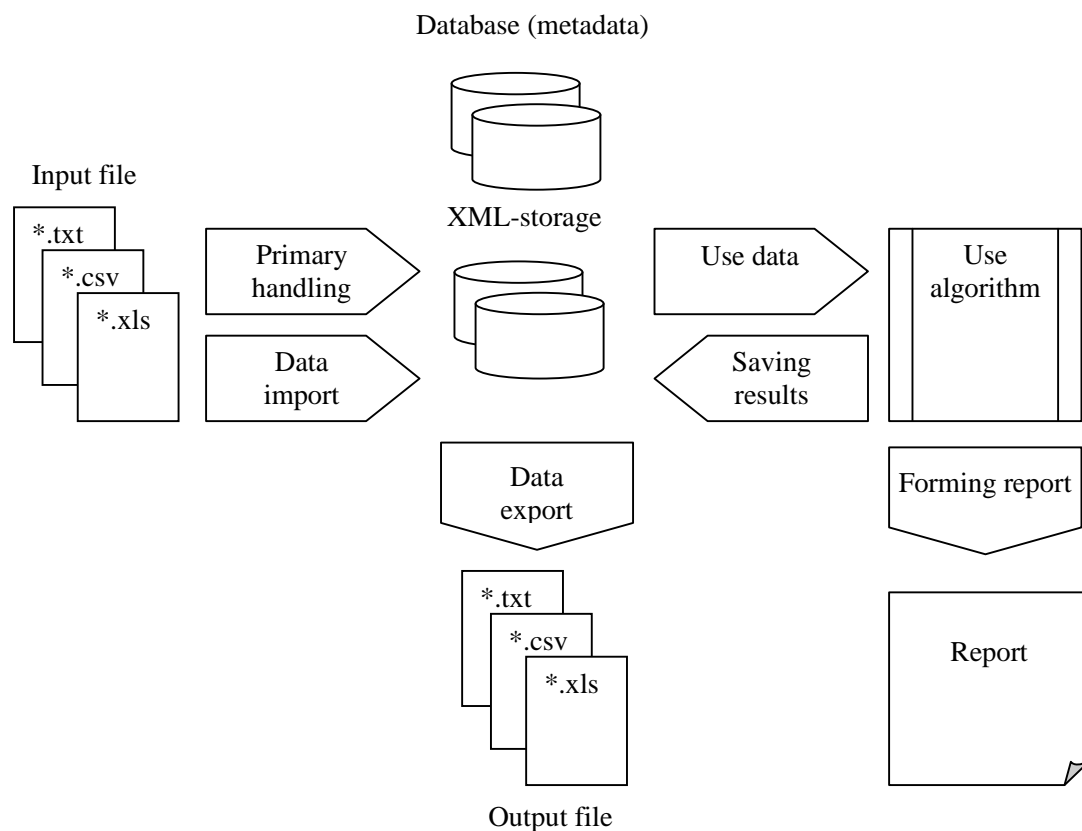


Fig. 1. Architecture of the integrated environment of handling and storing information

For forming files names which are contained in a storage is used MD5-key. Such decision allows to watch originality of document concluded in a storage. The process of verification consists in the following: at sending of document in a storage his name formed on the algorithm of MD5, at the receipt of document from a storage it follows repeatedly to get MD5-digest of this document, compare him to the current name and make conclusions about originality of document. In those case if the name and got on the algorithm of MD5 of digest coincide - a document has an original structure, in other case there is probability that in a storage now is unoriginal document.

4 Basic modules of the system

An export and an import data is executed on a scheme with a configuration file, which allows describing the parameters of data transformation. Not less important there is a conducting of a protocol file of data handling, which will allow on the stage of primary handling to find out the errors of data handling.

The module of import allows to execute data import from the different types of files for example *.txt, *.csv, *.xls to the XML-storage, using information that is contained in a configuration file. Reading information of input file the module of import works on principle of parser, and forms a XML-document. It is necessary to note that the module of import, forms in parallel a file of the transformed data, filling omissions, changing large or small data anomalous, and keeps, both an input and transformed file. Protocol of errors and executed actions is also forming.

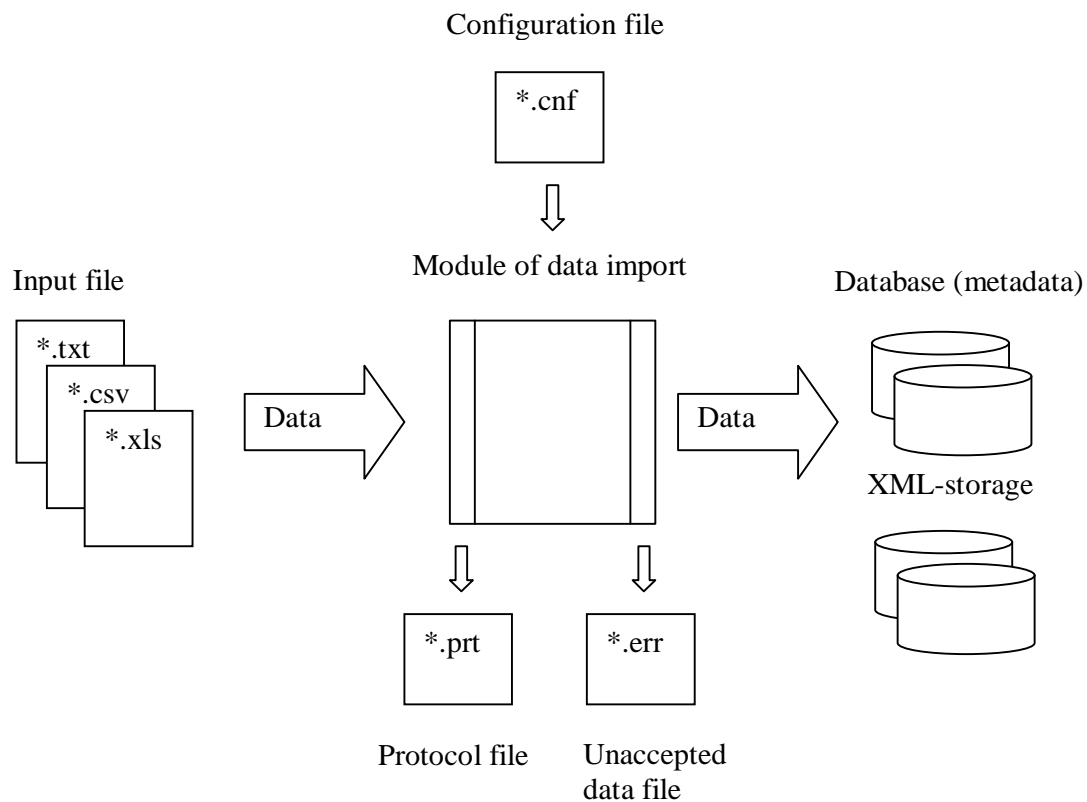


Fig. 2. Scheme of the module of data import

The module of export allows to execute the export of data for the use in other realization of algorithms. The data export like an import is executed by scheme with a configuration file. The module works as follows: an user chooses a necessary him file, coming from information that is contained in a database there is a place of finding of file in the file system, and the operation of transformation of XML-file is executed to the necessary format. It is thus possible to say that parser works in reverse direction.

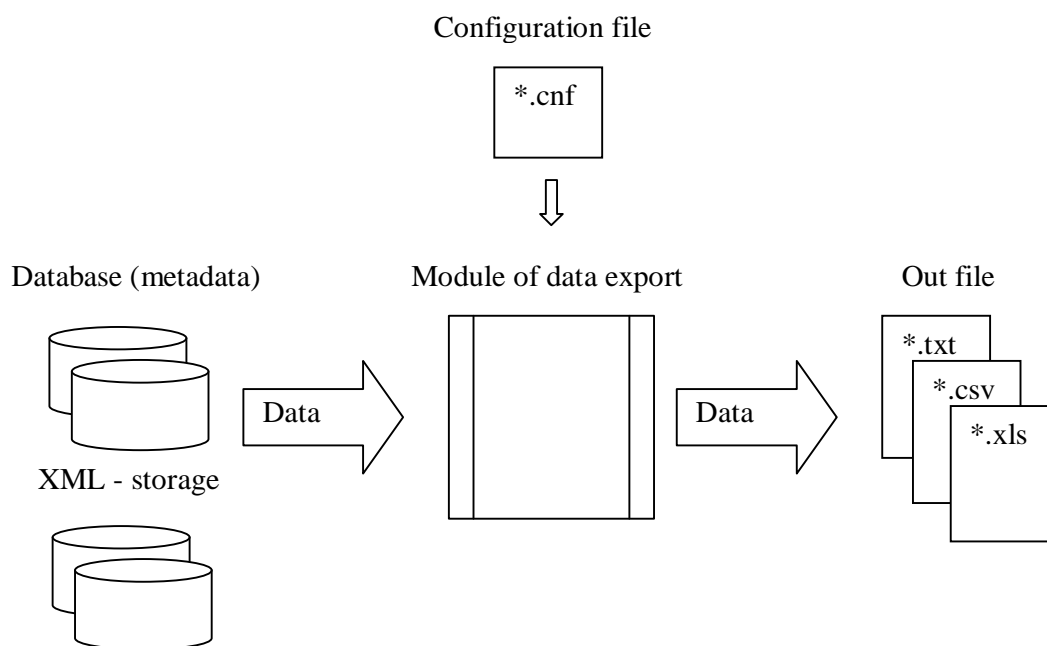


Fig. 3. Scheme of the module of data export

5 Conclusions

Architecture of an integrated environment is offered for information storage in the tasks of inductive modeling which makes it possible to freely manipulate with the available information. A model of the environment is constructed consisting of a relational database which contains only metadata and XML-storage of input statistical data and saved results of calculations.

References

- [1] Ivakhnenko A. G., Stepashko V.S.: Pomekhoustoichivost modelirovaniya (Noise-immunity of modeling). Kiev: Naukova Dumka, 216 p, 1985. (In Russian)
- [2] Ivakhnenko A. G.: Induktivnyi metod samoorganizatsii modeley slozhnykh system (Inductive method for self-organization of models of complex systems). Kiev: Naukova Dumka, 216 p, 1982. (In Russian)
- [3] Graves M.: Designing XML Databases. M: Publishing house of «Vil'yams», 640 p, 2002.
- [4] Thomas K., Karely B.: Databases. Design, realization and accompaniment. Theory and practice. – M.: William, 1440 p, 2003.
- [5] Christopher J. Date: Introduction to databases systems. — K.:BHV, 608 p, 1998.