

REZYUME. Maqolada nurlanishning ko‘rilayotgan jarayonlarga chiziqli va nochiziqli yutilishlarining ta‘sirini tahlil qilish imkoniyati yo‘q. Shunga qaramasdan, asosiy omillarning aytib o‘tilgan kompleks muhitda to‘lqinlarning nochiziqli o‘zaro ta‘siri hisoblashlarda, eksperimental tadqiqotlarda qanchalik murakkab masala ekanligi aniqlandi.

РЕЗЮМЕ. В статье нет возможности проанализировать влияние линейного и нелинейного поглощения излучения на рассматриваемые процессы. Тем не менее, было установлено, насколько сложным является вычисление и экспериментальное исследование нелинейного волнового взаимодействия основных факторов в упомянутой комплексной среде.

SUMMARY. In this article, there is no opportunity to analyse the influence of linear and nonlinear absorption of radiation on the processes under consideration. Nevertheless, it was established how complex it is to calculate and experimentally investigate the nonlinear wave interaction of the main factors in the mentioned complex medium.

METHODS FOR INTEGRATING INFORMATION EXCHANGE BETWEEN CORPORATE SYSTEMS

O.Askaraliev – *PhD on technical sciences*

Sh.Madaminov – *researcher*

Tashkent university of information technologies named after of Mukhammad al-Kharezmiy

Tayanch so‘zlar: integratsiya, avtomatlashtirilgan axborot tizimi, universal ma’lumotlar analizatori, qaror qabul qilish moduli, ma’lumotlar integratsiyasiga yondashuvlar.

Ключевые слова: интеграция, автоматизированная информационная система, универсальный анализатор данных, модуль принятия решений, подходы к интеграции данных.

Key words: integration, automated information system, universal data analyzer, decision-making module, approaches to data integration.

Indroduction. In the modern world, one of the complex tasks in the field of information technology is the integration of systems. Sooner or later, companies will face the task of automating business processes. At the same time, the integration of information systems is divided into two parts: application integration and data integration.

Data integration in information systems should be understood as the development and support of an interface for working with data from various independent heterogeneous sources as a single data model. The source of data can be various database systems, structured data files, websites, etc. With the increase in the volume and need for sharing data, the role of integration increases.

The process of data integration is relevant both in commercial tasks and in scientific ones (to combine research results from various sources).

This paper discusses methods of integrating information systems, architectures of integration systems, mechanisms for displaying data models, and modern approaches to data integration.

Integration of information systems. Integration of information systems means the creation of a single information space of the organization, which can be achieved by combining all implemented and planned for implementation automated systems into one integrated system. The problem of integration of automated systems arises when analyzing the following areas:

- development of technologies for the possibility of producing a large number of types of products, adjustment of production tasks;
- the need to organize the system of functioning of the organization based on flexible methods and approaches;
- elimination of barriers to the exchange of information within the organization;
- increasing manageability due to organizational measures by ensuring the transparency of information flows, operational management, and making coordinated decisions;
- development of a universal unified information space for all automated systems of the organization, which

means the potential possibility of real-time interchange between different components and modules.

The following areas of integration can be distinguished: software, functional, information, technical and organizational. Software is necessary to ensure the compatibility of the functional features of software tools that are used to solve specific deterministic problems.

The functional one combines the basic elements of the information and technical infrastructure, including components, for each of which a methodology for calculating the efficiency criterion, a basic functional and analytical model, information and functional relationships between the system modules are developed. As a result, a consolidated space of local areas of operation of all modules is provided.

The information one looks at the universalization of approaches to storing, collecting, using and presenting data at various levels of the management system. In this direction, the interconnected movement of information flows between the system modules is ensured.

The technical one is necessary to provide unified hardware and computing equipment and local networks, which in their totality are the basic elements for absolutely all areas of integration. The organizational one is necessary to support and maintain the activities of personnel in the field of management.

It is necessary to constantly keep in mind the undeniable fact that there is a wide variety of automated systems. Then it becomes obvious that the integration tasks can be very diverse. Problem statement. Due to the growing needs of the organization to obtain data from external sources, it becomes necessary to automate the data integration process. To do this, it is necessary to develop an architecture for the interaction of systems, thanks to which other IS will be able to easily and safely obtain data from external sources.

Often, third-party services for one reason or another do not want to open their APIs, in such cases, you have to extract the information yourself, hence the problems that a universal data collection system solves:

1. Extracting partially structured text information from sites.

2. Most sites with data are designed in such a way that in order to gain access to information, it is necessary to emulate user behavior, because some sites use special filters to provide robots with other information, i.e. the analyzer must unload information while remaining invisible to the site.

3. Bringing the data format specific to sites to an intermediate, similar (but not entirely identical) one. In addition, it is necessary to track changes in the data format and notify about it.

4. Tracking obvious duplicates, updating changed information.

5. Parsing text information from websites and converting it into a fully structured format.

6. The required architecture is necessary for a system that automates the processes of detecting collateral fraud and verifying (evaluating) real estate/mortgage portfolios accepted as collateral (anti-fraud system).

To identify signs of fraud (external and internal), the system analyzes both previously collected data and current market data on collateral. Based on the characteristics of the object, analogs are selected on the real estate market and the value of the object is calculated based on their prices. The system is used both for evaluating individual objects and for evaluating and analyzing a portfolio of objects.

The software part of the required system collects information about offers for sale of residential real estate. This program must perform the following tasks:

- internet sites and download HTML pages open for public use. The number of sites used is 150. The volume of daily downloaded ads is about 32 gigabytes;
- converting read ads, which have a specific format for each site, into some intermediate unified representation (set and content of fields);
- tracking changes in the document description format on the site. (It is assumed that the structure of the site pages changes on average 1-2 times a year, then when downloading information from 150 sites, it should be expected that the frequency of site revision can be daily);
- tracking obvious duplicates of ads, as well as accounting for updates to information in the ad;
- calculating the ad exposure period on the site;
- saving the ad archive (HTML + Page image in graphic format) in the primary DB with the ability to search by ID, URL;
- the system provides a number of algorithms that can be built into its various parts to solve internal problems (for example, address parsing, duplicate detection), which can extend across many subsystems (calculation of the ad exposure period).

Unstructured data received from external sources must be stored in an ordered format.

Approach to solving the task. As organizations develop, their internal architecture, business processes, user interfaces change, processes accelerate, and in such a situation, the integration task becomes a serious problem. As they continue to develop, companies become large, tasks become complex, and distribution appears, which requires support. In addition, many organizations, depending on the budget and tasks, use different platforms and tools from different manufacturers (heterogeneity), which

must be supported in the future. Also, do not forget about obsolete technologies, systems or informal structural data, which also have to be used. Integration tasks are not limited to the organization, it is increasingly necessary to receive data from partners, clients, contractors, government agencies and other external sources (intersystem integration), etc.

Parameters responsible for the complexity of integration:

1. Conceptual difference - the basis of this difference is that the developers of different IS have made conceptually different assumptions, so these systems cannot be correctly combined. This problem is solved by introducing another abstraction layer that will not contradict conceptually both systems. In this case, two implementation options can be distinguished:

- the resulting system is centralized, and all other integrated systems are transformed into subsystems;
- using the architecture of a broker, an intermediary, which is not a center, but only provides a layer between the systems. All integrated systems remain independent.

2. Technological difference – based on the incompatibility of data exchange formats, interaction protocols and interfaces. It is solved by using brokers, layers, envelopes and other not entirely beautiful, but quite reliable means.

The general problem looks like this: it is necessary to integrate N information systems, taking into account these factors, with the minimization of the number of layers. With the first solution to the problem, it turns out that between N systems there will be $\frac{N(N-1)}{2}$ connections with two-way interaction of $N \times (N - 1)$ interfaces.

There are many solutions to this problem, but the following classic methods have been identified:

- data-level integration - the idea behind this method is that a certain array of applications sends a request to a single database or to several different databases that are linked by replications;
- broker-level integration - the essence of this method is to develop and implement an additional (not the main) software module that accesses multiple systems;
- standardization - a typical classic method that requires the use of various standards (both international and state, as well as industry-specific) when developing systems for backward compatibility;
- user-level integration - this method is undesirable, one might say, it is not automation. The idea behind the presented method is that users exchange data between systems in a non-systemic way, but in other available ways. For example, via mail, copying data, transferring files, etc. This method is used as a last resort, when it is impossible to freeze a business process during an update or modernization;

– integration at the service level – this method contains a very elegant concept of integration, in which interfaces and data formats are fixed from all sides. When using this integration method, it is very easy to work out and configure intersystem logic.

Comparison of integration methods is presented in Table 1.

Table 1. Comparison of integration methods

Methods	Advantages	Flaws
Stantadtization	Compatibility	Long term introduction of new standards if necessary
Integration at the broker level	Versatility	Complexity, labor intensity, high cost of development
Data level integration	Low cost of integration	Different applications may leave data in inconsistent states; code duplication
Integration at the service level	Rapid development of intercorporate business logic	There is fixation
Integration at the users level	Used as a last resort	Non-automated integration

Selecting an approach to integration and development tools. Data integration will be performed not only at the physical, but also at the logical and semantic levels. The data itself is important for this development, as well as its logical connection in sources, taking into account the nature of this information.

Therefore, based on the analysis of the tasks to be solved and the requirements for them, it was decided to automate the process of integrating unstructured data by developing an ETL system that will allow implementing a flexible, easily supported, expandable solution based on open source technologies.

The presented system will have its own database for consolidating information from target sources, tracking changes and transferring data to other IS and applications.

Architecture of the developed system. After selecting the approach to integration and development tools, the architecture of the data collection system was proposed. The entire architecture (Fig. 1) can be divided into two parts.

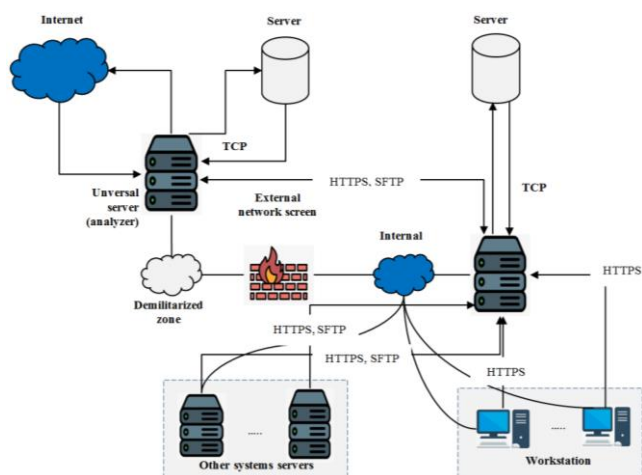


Fig. 1. Data Collection System Architecture

The first part contains an application server that interacts with external sources and a database server to store the collected data. The entire part is located in a DMZ (demilitarized zone).

In computer security, a DMZ, or demilitarized zone (sometimes called a perimeter network), is a physical or

logical subnet that contains and provides an organization's external services to an untrusted network. Typically, this is a larger network such as the Internet. The purpose of a DMZ is to add an extra layer of security to an organization's local area network (LAN): an external network node can only access what is exposed in the DMZ, and the rest of the organization's network is a firewall. The DMZ functions as a small, isolated network located between the Internet and the private network, and if its design is effective, allows the organization to spend extra time detecting and remediating breaches before they penetrate internal networks. The name comes from the term "demilitarized zone" - an area between nation states in which military operations are not permitted.

From workstations, via the application server in the second part, there is access to the universal analyzer server via an interface for specifying external sources that will be analyzed.

Description of the software module. The PERL language was chosen for developing the analyzer, because the main feature of the language is its rich capabilities for working with text, including working with regular expressions built into the syntax.

The following modules were used in developing the application:

- CGI module, which is designed to simplify the creation of HTML documents;
- DBI module, necessary for working with the database;
- LWP module library, which allows you to fully use the capabilities of http(s) protocols.

The application is written in such a way that the operator enters some rules, conditions, settings and templates (which are stored in the application database) through the interface. Then, depending on the settings, the analysis of online services, selection of information and downloading to the database begins.

The universal analyzer solves all the tasks:

- extraction of partially structured text information from sites;
- while receiving information, the analyzer is able to evaluate the encoding of the received data and lead to the encoding of the target table for correct display of data. If the operator can bypass this error, then the robot will be able to do it based on the data entered by the operator;
- checking the adequacy of the received data, notifying the user about problems - upon receipt of a clearly short message, in case of an unsuccessful attempt to split the page into a structured format, the user will be notified and options for solving the problem will be offered;
- the universal analyzer does not require a specially trained person to use it. Any PC user can easily handle all the settings and create a task for data unloading. This robot has many settings that allow the user to use it for various sources of information with different data structures without going into the source code;
- this analyzer has an open API with basic methods so that it can be used without a standard interface, and so that other developers can write their own interface, taking into account the features of their systems;
- in addition to directly receiving structured data from the database, they can be received as a database dump, an excel table in an email, or downloaded directly.

Fig. 2 shows the scheme of extraction from external sources, transformation and loading of data into the storage by the anti-fraud system.

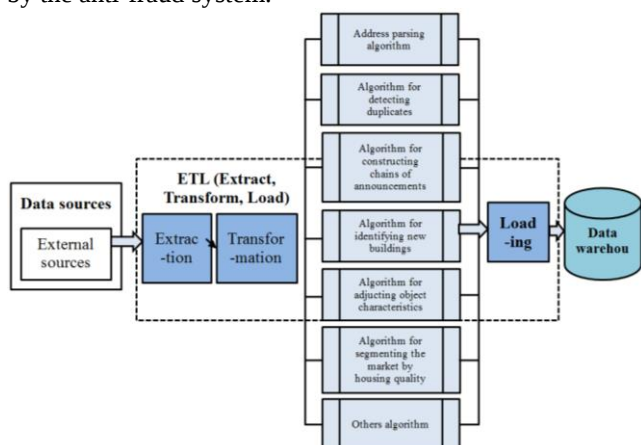


Fig. 2. Data extraction scheme from external sources

Results of system implementation. The presented development has increased the efficiency of fraud detection both on the part of clients and on the part of employees processing loans, and has also made it possible to evaluate real estate/mortgage portfolios accepted as collateral online and accelerated the decision-making mechanism on issuing a mortgage.

Before the implementation of the system with a universal analyzer, two employees (an operator and a verifier) were required for a full working day (working 6 hours 4 days a week). All applications were processed manually, real estate websites (about 100 pcs.) and a database prepared by the organization were analyzed. The productivity of these employees was limited to processing 320 applications per day (~1 application in 3 minutes). During the work, 6 % of errors occurred due to input errors.

After the introduction of the presented system, productivity increased. The error rate has decreased to 0.007%. Automatic duplicate control and history maintenance are performed. Using the example of an anti-fraud system developed on the basis of a universal analyzer, one can see how a universal analyzer allows automating processes related to data integration.

Conclusion. The presented article considered various problems that arise in the process of integrating information systems. It also revealed approaches to solving the problem, including planning the system interaction architecture and developing a universal data analyzer. In general, unified approaches to integrating information systems were put forward.

References

- Schmarzo B. Big Data: Understanding How Data Powers Big Business / Bill Schmarzo. Hoboken, NJ : Wiley, 2013. –P. 240.
- Saidov A., Mirboboev M., Almetov Sh., Ganieva N., Boboqulov I. “Bojxona organlarining axborot tizimlari asoslari” – I tom. –Toshkent: 2016. 421-b.
- Askaraliev O.U., Malikova N.T., Naim N.A., Sharma Sh., Chohan Zh., Kumar R., Singh S. Intelligent Decision Support Systems in Integrated Management Systems – A Novel Multicriteria Computational Control System. Advancements in Civil Engineering & Technology, AIP Conference Proceedings 2558, 020001 (2023). 03 February 2023. <https://doi.org/10.1063/5.0121301>.
- Осипов Д.С., Грошев Ф.В. О скорости передачи данных в одной системе множественного доступа. // Информационно-управляющие системы. 2012. № 1. -С. 67-75.
- Vikulov E.O., Denisov O.V., Denisova L.A. Data distribution system preparation of server stations data Journal of Physics: Conference Series, 2018.
- NGINX: Advanced Load Balancer, Web Server, & Reverse Proxy [Электронный ресурс] URL: <https://www.nginx.com> (дата обращения: 23.09.2023).
- Hastie T., Tibshirani R., Friedman J. H. The elements of statistical learning: Data mining, inference, and prediction. -New York: Springer Verlag, 2001.
- MATLAB Makers of Matlab and Simulink. [Электронный ресурс] URL: <http://www.mathworks.com> (дата обращения: 12.11.2023).
- Ghomi E.J., Rahmani A.M., Qader N.N. Load-balancing algorithms in cloud computing: A survey // Journal of Network and Computer Applications. 2017.

REZYUME. Fan va texnika sohasida integratsiyalashgan o‘zaro axborot almashinuvini tashkil etishning yagona standartlashtirilgan algoritmi hozirgacha ishlab chiqilmagan. Ushbu maqolada integratsiya masalalari tahlil qilinadi va o‘rganadi. Jarayonlar davomida ba’zi integratsiyalashning qonuniyatlari aniqlandi va ko‘rsatilgan masalalarni yechish uchun yondashuv tanlandi. Ushbu yondashuv turli tizimlar o‘rtasida o‘zaro ma’lumot almashish imkoniyatini yaratdi.

РЕЗЮМЕ. В области науки и техники до настоящего времени не разработан единый стандартизированный алгоритм организации интегрированного взаимного обмена информацией. В данной статье анализируются и изучаются вопросы интеграции. В процессе были выявлены некоторые закономерности и выбран подход к решению обозначенных вопросов. Данный подход создал возможность для взаимного обмена информацией между различными системами.

SUMMARY. In the field of science and technology, a single standardized algorithm for organizing integrated mutual information exchange has not been developed to date. This article analyzes and studies the issues of integration. In the process, some patterns were identified and an approach was chosen to solve the indicated issues. This approach created an opportunity for mutual information exchange between different systems.