## МЕТОДИ ТА СИСТЕМИ ОПТИКО-ЕЛЕКТРОННОЇ І ЦИФРОВОЇ ОБРОБКИ ЗОБРАЖЕНЬ ТА СИГНАЛІВ

**M.L. KOVALCHUK, P. LUCIN, M.P. GORSKY, I.V. SOLTYS**

# DESIGN AND CREATION OF AN INFORMATION SYSTEM FOR ANALYTICAL DATA PROCESSING

*Yuriy Fedkovych Chernivtsi National University,*
*Kotsyubynsky 2, 58012 Chernivtsi, Ukraine, e-mail: m.kovalchuk@chnu.edu.ua*

**Анотація**. Розроблено інформаційну систему, що дає можливість визначити взаємозв'язок соціально-економічних факторів розвитку країни засобами множинного кореляційного аналізу. Ефективний аналіз показників може значною мірою допомогти управлінцям приймати правильні рішення, що позитивно вплинуть на розвиток країни. Програмний продукт реалізовано у вигляді веб-сайту. У процесі розробки використовувались наступні технології та фреймворки: Java, Java EE, Spring framework, Spring data, Spring security, MySQL, Maven, Git, jUnit, Mockito, Liquibase, Tomcat, JSP.
**Ключові слова**: інформаційна система, кореляційний аналіз, інтелектуальний аналіз даних.

**Abstract**. An information system has been developed that makes it possible to determine the relationship between socio-economic factors of the country's development by means of multiple correlation analysis. Effective analysis of indicators can greatly help managers to make the right decisions that will positively affect the development of the country. The software product is implemented as a website. The following technologies and frameworks were used in the development process: Java, Java EE, Spring framework, Spring data, Spring security, MySQL, Maven, Git, jUnit, Mockito, Liquibase, Tomcat, JSP.
**Key words:** information system, correlation analysis, data mining.

## INTRODUCTION

The use of modern technologies greatly simplifies the accumulation of statistical data relating to various branches of the country's economy. Effective analysis of indicators can greatly help managers to make the right decisions that will positively affect the welfare of citizens. By accumulating statistical data for a certain period of time, it is possible to obtain information that will allow not only to see the level of socio-economic development of the country, but also the mutual influence of seemingly unrelated factors. Therefore, the authors of this work decided to develop an information system that would allow, through multiple correlation analysis, to determine the influence of independent factors on the chosen one.

During the development of this system, such tasks as the choice of system architecture were solved; technology stack selection; database design; system design; implementation of the security level and mechanism of multiple correlation analysis; formation of business logic, namely data processing mechanisms and business rules; forming the user interface [1, 2, 3].

The developed system is focused on a specific subject area, which allows taking into account its specificity and applying specialized mechanisms for data analysis, in particular - a mechanism for automatically forming a conclusion to the conducted research according to the received information. The practical significance of the developed system is that it can provide the opportunity to analyze the interaction of factors of socio-economic development, which, in turn, will allow making correct and effective management decisions to ensure rapid and productive growth [4, 5, 6].

## 1. FORMULATION OF THE PROBLEM

Analyzing already existing systems, it is possible to single out a large number of really high-quality software products that provide powerful functionality for processing any data, presenting it in the form of various graphs, diagrams , etc. However, most of the analyzed systems have a number of shortcomings [7, 8].

These include the need to install a software product; the need to purchase a license key; abstraction from a specific type of data; the absence of the possibility of automatically forming a certain conclusion regarding the conducted research. The developed system takes into account and eliminates the specified shortcomings. Since it is web-oriented, there is no need to install third-party software products, purchase license keys, and configure third-party plugins . In addition, the above-mentioned system is focused on a specific field of application, namely the analysis of socio-economic development monitoring data, therefore it has mechanisms for data processing taking into account the specifics of the subject area, as well as, most importantly, means for automatically forming a certain conclusion based on the results of the study [9, 10, 11].

Web-orientation will also avoid many problems, including the chaotic use of a software product. With the desktop approach, the end user receives a distribution file that installs the software product on his personal computer, or a package of files necessary for the application to work. But this is a vulnerability, because by the means and methods of decompilation and disassembly, unscrupulous third parties can obtain the source code of the program and modify it for a beneficial purpose. Web-orientation allows to reduce the risk of the above problem, because all the software code is stored on the server and the end user does not have access to it.

Very often, pair correlation models are used to analyze the relationship between factors. However, in practice, it is often necessary to study phenomena that occur under the influence of many different factors, where the influence of each of them separately is not too significant and does not show the overall picture of the interaction process. At the same time, the total effect turns out to be quite strong and allows us to draw the necessary conclusions. That is why we used multiple correlation, when one dependent variable is considered as a function of several independent ones.

It should also be taken into account that some methods of processing multidimensional statistical data require prior normalization. It consists in bringing the data into compliance with the new form of presentation. Such transformations make it possible to exclude the impact on the results of the analysis of the adopted measurement units.

The peculiarity of indicators of economic and social development is that a strict and unified list of quantitative parameters that clearly represent the empirical socio-economic system has not yet been developed. Also, an exhaustive list of operations that must be carried out in order to evaluate one or another determining factor has not been established. [1, 12]

Short series of observations and the not always experimental nature of the data greatly complicate the process of registering indicators and often call into question the scientific significance of the results of their measurements. Due to the colossal spatio-temporal variability of objects, there is no certainty that the available samples reflect real processes. The subjective factor plays a very important role: socio-economic values are formed in the course of a certain activity and somehow characterize this activity. [1, 13]

Indicators loaded into database tables are presented in a wide variety of measurement scales: nominal, ordinal, and metric. Indicators measured in metric scales have a wide variety of units of measurement, scale, reference points, and variation intervals.

Empirical series of measurements obey the most diverse laws of distribution, far from theoretical normal or uniform. Graphs of dependencies often look like stochastic fluctuations approaching "white noise". The problem of identifying "outliers", filtering anomalous and restoring missing values takes on threatening proportions. In this regard, the development of an advanced input data preprocessing system is crucial for obtaining adequate results of mathematical modeling [2, 14].

In statistical data processing, normalization is widely used, that is, linear transformation of all attribute values in such a way that the attribute values fall into intervals comparable to the value:

$$x_{ij} = \frac{x_{ij} - A}{B},$$

where $x_{ij}$ is the j-th coordinate of the i-th vector, A and B are some predetermined numbers, which we will call characteristic scales.

These numbers can be determined based on the statistical characteristics of the distribution of empirical samples (normalization according to statistics), or given according to some a priori considerations (normalization according to standards). The background or critical values of the indicator, the best and worst "favorable" values, and other estimates lexically related to the problem of analyzing critical or permissible loads can act as a

"standard". These estimates are easily perceived, but there are no methods of their correct calculation, and existing individual attempts at economic normalization should be considered subjective. In the multidimensional data space, there are several scales of normalization by statistics, when the variation series of each selected indicator is transformed using sample statistical characteristics. First, it is the geometric center of the multidimensional space of data points $\overline{X}$(that is, the average value of all features), the square root of the total variance $\sigma$ , called the root mean square deviation, and the scale R, which characterizes the maximum spread in the data cloud. [3]

Normalization of all features by R leads to the fact that the entire data space is placed in a sphere of unit radius, and the corresponding preprocessing formula has the form:

$$\widetilde{X_\iota} = \frac{X_i - X}{R},$$

where $X_i$ and $\widetilde{X_\iota}$ are new and old vectors of feature values.

However, let's return to practical solutions. Let us introduce such a concept as the normalized scale (NS) of an indicator, which is characterized by the following properties: for all indicators converted into NS, a single range of the area of existence is established, varying from B min to B max ; the distribution of the variation series of the indicator on the NS scale corresponds to the principle of the maximum entropy of each of the input variables.

As part of the study, an ordinal scale was adopted as a normalized scale, in which B min = 1, and B max = K b , where K b is the dimension of the scale (number of gradations). For reasons of unification, for most of the original indicators measured in metric scales, K b was taken equal to 6. The dimensionality of K b for indicators originally measured in ordinal or nominal scales was chosen each time based on the specifics of normalized data. [4]

## 2. STACK OF USED TECHNOLOGIES

In the process of developing the software product, the following technologies and frameworks were used : Java , Java EE, Spring framework , Spring data , Spring security , MySQL , Maven , Git , jUnit , Mockito , Liquibase , Tomcat , JSP.

The project is built according to the MVC design pattern. The use of this model is due to the speed of development with a small number of pages, which increases the productivity of work at the initial stage.

The main frameworks in this project were Spring for the development of the main functionality and Hibernate for working with the database.

JSP technology was used to design the website pages. This choice is due to the speed of development with a small number of pages and compliance with the principles of encapsulation, modularity and weak connectivity of components.

Also, for a more detailed description of the application architecture, an interpretation of the MVC design template should be provided (Fig. 1).
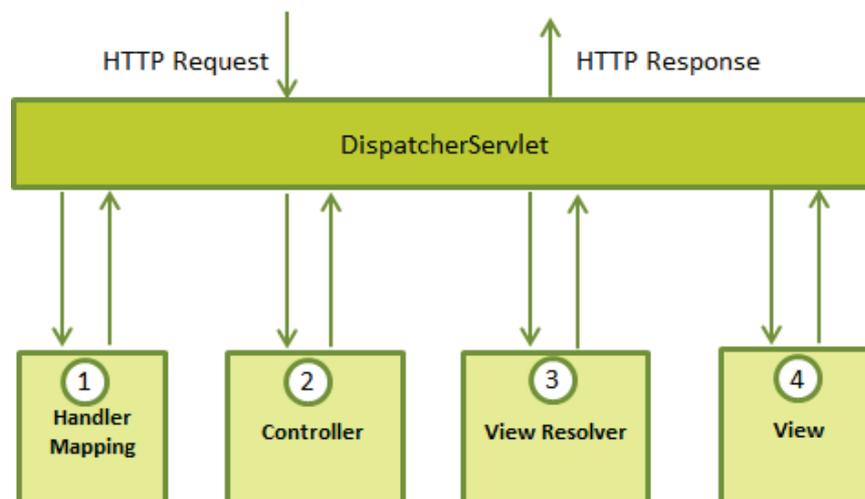


Figure 1 – Interpretation of the MVC design pattern

Client requests are sent to the serviette manager or the so-called front controller , which forwards these requests to the appropriate controllers. After that, the corresponding controller processes the information and gives a response, which is processed by the View component Resolver , which aims to deliver the appropriate page to the client after the View Resolver will work. Next, the user is given a page with the appropriate content that was defined in the controller.

## 3. SYSTEM DEVELOPMENT

Considering the orientation of the project on the client-server architecture, the Java platform was used Enterprise Edition (J2EE), as well as Spring Framework for building client-server architecture and software product architecture. Automatic project assembly tools, namely Apache , were used to create a single system from separate modules Maven . The Jakson JSON library was used to ensure the client-server mechanics of data exchange . This system operates using the Tomcat web server , although a built-in servlet container , such as Jetty , can be used for this purpose .

The first step in the development of this system was the analysis of the subject area. The source of statistical data was the website of the State Statistics Service of Ukraine, which provides information on data on the socio-economic development of the country for each month, as well as an archive of data for recent years.

To realize the possibility of using the software product on the server, it was necessary to create a so-called web module from the project files, it can be a JAR file or a WAR file.

When designing, it was decided to use a WAR file. This decision is due to the following advantages of this method of packaging: ease of development, testing and deployment; the version of the deployed application is easy to identify; you can sign WAR files using standard Java utilities included in the J2SE package; all JEE containers support WAR files.

When creating a web module, it is necessary to pay attention to its typical structure, which contains two main directories, which contain project resources and test resources, respectively. The directory containing the project resources consists of three subfolders containing the program code, static resources (for example, the configuration files of the logging system or the general configuration of the project) and a folder for web resources, namely pages, CSS style configuration files ( Cascading Style Sheets ) and the web.xml file, also called the deployment descriptor .

The next step in the development of this system was the design of the database, the formation of tables and connections between them.

frameworks were selected . Apache is used to link all system components into a single entity Maven , for this, XML configuration files and the object model of the project, which is stored in a specialized "pom.xml" file, are used. This file is constantly updated by the development environment when the corresponding item is enabled in its configuration.

Properties " groupId ", " artifactId ", " version " are responsible for the identification of components, there are no two identical triples of these properties, this is done to avoid confusion and duplication of components. When modifying code or simply updating, Apache Maven checks the dependency lists and downloads components that are missing from the project, or removes redundant ones. The download is made according to the list of repositories Maven , which is configurable in the development environment: you can add public repositories Maven , local and special remote counterparts that contain test beta versions of components.

After designing the object model of the project, work was started on the level of access to the database. Since the project uses Spring technology data JPA, then the development of this level is quite simple: a new interface is created, which is inherited from JpaRepository <T,K>. The JpaRepository interface already contains certain methods that should be in every class that is used to access the database, namely: save, find all, delete , etc. Advantages of using Spring data are quite significant, namely: there is no need to write bulky, heavy and unreadable classes with a lot of code that does not even relate to working with the database, but is responsible for error handling , logging , etc. In the end, we get a class in which out of 30 lines of code in one method, only 2 lines are responsible for business logic.

Spring data "on the fly" generates classes from the interfaces we created. No need to write native queries or any cumbersome constructions like opening/closing sessions and transactions. The entire request can be written with only the method name and arguments, using special expressions and combinations, which are described in the Spring documentation . This approach allows you to abstract from a specific database and allows you to freely migrate a project from, for example, PostgreSQL to Neo4J.

The next stage of work was the development of so-called "contracts" for the level of services. This is necessary to comply with the IoC principle : the service is integrated into the component, not its implementation, such a mechanism allows the developer to easily change the implementation without changing the business logic, and also provides an opportunity to abstract from the specific implementation of the service and

concentrate on the development of another class.

After the "contracts" were developed, the services themselves were made. With the help of Spring annotations, the DI level of access to databases in the implementation of services was performed, and the basic business logic was also implemented at the level of services.

The next step requires parallel development of controllers and JSP pages. The work at this step is carried out step by step. For each page, a corresponding controller or a method in the controller is created (depending on its semantics), and markup is developed using special directives.

At this stage of work, the following parts of the software product are ready: database, database access level, service level (contracts and their implementation), controllers and pages. This could be the end of the job, but the task is to connect Spring security for user authentication and application security.

Next, it is necessary to make an implementation at the service level for the org.springframework.security.core.userdetails.UserDetailsService interface, which is responsible for loading users, after which it is necessary to configure Spring security and add its objects to the general application context.

## 4. ANALYSIS OF MONITORING DATA

With the help of the developed software product, a number of studies of the monitoring data of socio-economic development were conducted.

As an example, in one of the conducted studies, "Export of goods and services" was used as a dependent (researched) factor. "Nominal average monthly salary of one employee", "Volume of realized industrial products", "Volume of completed construction works", "Volume of agricultural products" were chosen as independent factors.

As a result of the study, a multiple correlation coefficient equal to 0.7562456 was obtained, which indicates a fairly strong connection between the above-mentioned factors and indicates the degree of their influence on the subject. Based on the received information, management decisions should be made for the growth of the above-mentioned independent factors, which in turn will lead to an increase in the export of goods and services, and the growth of this indicator will increase the supply of goods and services, their quality, and therefore competitiveness in general.

## CONCLUSION

As a result of the implementation of this project, an information system was created for the analytical processing of social and economic development monitoring data.

Thanks to the finding of multiple correlations between input data, the system allows you to detect seemingly insignificant relationships between input factors and provide the user with the information necessary to make correct management decisions.

A number of tools were used to implement the task, including the Java programming language , the JEE platform, and Spring framework, Spring data , Spring security, MySQL database, Maven automated build system, Git version control system, jUnit testing tools, Mockito, Liquibase database migration system , servlet container Apache Tomcat.

## REFERENCES

1. I.I. Eliseeva. "Practical work on econometrics: Study guide." / Eliseeva I.I. M.: Finances and Statistics, 2005 - 192 p.
2. T.G. Maksimova. "Basic econometrics: Study guide" / Maksimova T.G., Verzlin D.N. - St. Petersburg: TEI, 2009. - 80 p.
3. Yu. V. Kozhevnikov, "Probability Theory and Mathematical Statistics." / Yu. V. Kozhevnikov - M.: Mashinostroenie, 2002.
4. N. Sh. Kremer. "Probability theory and mathematical statistics. A textbook for universities." / N. Sh. Kremer - M.: UNITY-DANA, 2000.
5. F.F. Vodovatov, A.A. Chelny, V.P. Veyko, M.N. Libenson Lasers in technology /. - M.: Energy, 1975. - 215 p.
6. O. V. Galochkin, V. M. Zhikharevych, et al. Effect of powerful millisecond laser radiation on the depth of the melted layer in CdTe and $Cd_{0.8}Mn_{0.2}Te$ crystals // Solid state physics and chemistry. – 2012. – Vol. 13, No. 1. - P. 224-229.

7.  Highly linear Microelectronic Sensors Signal Converters Based on Push-Pull Amplifier Circuits / edited by Waldemar Wojcik and Sergii Pavlov, Monograph, (2022) NR 181, Lublin, Comitet Inzynierii Srodowiska PAN, 283 Pages. ISBN 978-83-63714-80-2.
8.  Submicron and Nanoscale Structures of Electronics / [Z. Gotra, I. Grigorchak, S. Pavlov, etc.]. - Chernivtsi: Technological Center, 2014. – 839 p.
9.  Physical foundations of biomedical optics (Monograph) / [Pavlov S.V., Kozhemyako V.P., Kolisnyk P.F. KozlovskaT. I., Dumenko V. P.] – Vinnytsia: VNTU, 2010. – 155 p.
10. V.V. Kukharchuk, W. Wójcik, S.V. Pavlov, etc. "Features of the angular speed dynamic measurements with the use of an encoder", Informatyka, Automatyka, Pomiary w Gospodarce i Ochronie Srodowiska, 2022, 12(3), pp. 20–26.
11. Oleksandr I. Nikolskyy, Vladimir G. Krasilenko, Yosyp Y. Bilynsky, Anzhelika Starovier, "Using LabView for real-time monitoring and tracking of multiple biological objects", Proc. SPIE 10170, Health Monitoring of Structural and Biological Systems 2017, 101703H (5 April 2017).
12. O. D. Azarov, O. D. Dudnyk, M. Duk, D. Porubov, "Static and dynamic characteristics of the self-calibrating multibit ADC analog components", Proc. SPIE 8698, Optical Fibers and Their Applications 2012, 86980N (11 January 2013).
13. A. D. Azarov, A. I. Chernyak, P. A. Chernyak, "Class of numerical systems for pipeline bit sequential development of multiple optoelectronic data streams", Proc. SPIE 4425, Selected Papers from the International Conference on Optoelectronic Information Technologies, (12 June 2001).
14. M. A. Arbib, The Handbook of Brain Theory and Neural Networks, Second Edition, London, England: The MIT press, 2003.

**KOVALCHUK MYROSLAV –** Ph.D., Docent of Computer Science Department, Yuriy Fedkovych Chernivtsi National University, Chernivtsi, Ukraine, *e-mail: m.kovalchuk@chnu.edu.ua*

**LUCIN PAVLO –** Master of Computer Science, Computer Science Department, Yuriy Fedkovych Chernivtsi National University, Chernivtsi, Ukraine, *e-mail: lucin@chnu.edu.ua*

**GORSKY MYKHAILO** – Ph.D., assistant professor of Computer Science Department, Yuriy Fedkovych Chernivtsi National University, Chernivtsi, Ukraine, *e-mail: m.gorskiy@chnu.edu.ua*

**SOLTYS IRYNA** – Ph.D., Assistant Professor of Optics and Publishing Department, Yuriy Fedkovich Chernivtsi National University, Chernivtsi, Ukraine, *e-mail: i.soltys@chnu.edu.ua*