MINISTRY OF EDUCATION AND SCIENCE OF UKRAINE
SUMY STATE UNIVERSITY
UKRAINIAN FEDERATION OF INFORMATICS

PROCEEDINGS

OF THE V INTERNATIONAL SCIENTIFIC CONFERENCE

ADVANCED INFORMATION SYSTEMS AND TECHNOLOGIES

AIST-2017

(Sumy, May 17–19, 2017)

SUMY
SUMY STATE UNIVERSITY
2017

UDC 004(063)
A

A Advanced Information Systems and Technologies: proceedings of the V international scientific conference, Sumy, May 17–19 2017 / Edited by S. I. Protsenko, V. V. Shendryk – Sumy: Sumy State University, 2017 – 128 p.

This book comprises the proceedings of the V International Scientific Conference "Advanced Information Systems and Technologies, AIST-2017". The proceeding papers cover issues related to system analysis and modeling, project management, information system engineering, intelligent data processing computer networking and telecomunications. They will be useful for students, graduate students, researchers who interested in computer science.

UDC 004(063)

ISSN 2311-8504

International Scientific Committee:

A.N. Chornous, Sc.D (Ukraine)
A.S. Dovbysh, Sc.D (Ukraine)
O.A. Borisenko, Sc.D (Ukraine)
E.A. Lavrov, Sc.D (Ukraine)
V.O. Lyubchak, PhD (Ukraine)
S.I. Protsenko, Sc.D (Ukraine)
A.M. Kulish, Sc.D (Ukraine)
M.M. Glybovets, Sc.D (Ukraine)
Yu. I. Grytsyuk , Sc.D (Ukraine)
I.V. Grebennik, Sc.D (Ukraine)
O.O. Yemets, Sc.D (Ukraine)
D.D. Peleshko, Sc.D (Ukraine
V.I. Lytvynenko, Sc.D Ukraine)
K.V. Koshkin, Sc.D (Ukraine)
E.A. Druzynin, Sc.D (Ukraine)
S.I. Dotsenko, PhD (Ukrain)
T.V. Kovalyuk, PhD (Ukraine)
A. Pakštas, PhD (United Kingdom)
O.Romanko, PhD (Canada)
I. Polik, PhD (USA)
V.Kalashnikov, Sc.D (Mexico)
S. Berezyuk, PhD (Canada)
P. Davidsson, PhD (Sweden)
M. Biagi, PhD (Italy)

Organizing Committee:

Protsenko S. I., Sc.D, co-chairman (Ukraine);
Shendryk V. V., PhD, co-chairman(Ukraine);
Vaschenko S. M., PhD, co-chairman (Ukraine); Parfenenko Y. V., PhD (Ukraine), Nahornyi V. V. , PhD (Ukraine); Shulyma O. V. (Ukraine), Boiko A. O. (Ukraine), Zakharchenko V. P. (Ukraine), Tovkus O. I. (Ukraine), Shendryk S.O. (Ukraine), Rovna A.V., executive secretary (Ukraine).

Contacts:

Address: AIST conference, Sumy State University
2 Rimsky-Korsakov Str., Sumy, 40000, Ukraine
website: www.aist.sumdu.edu.ua
e-mail: aistorgcom@gmail.com.

CONTENTS

SESSION 1. SYSTEM ANALYSIS AND MODELING

SESSION 2. PROJECT MANAGEMENT

SESSION 3. E-LEARNING TECHNOLOGIES

SESSION 4. ENGINEERING OF INFORMATION SYSTEMS

SESSION 5. INTELLIGENT DATA PROCESSIN

SESSION 1
SYSTEM ANALYSIS AND MODELING

About properties of linear stochastic optimization problems on arrangements

O.O. Iemets, T.M. Barbolina
Poltava University of Economics and Trade, Poltava V.G. Korolenko National Pedagogical University, tm-b@ukr.net

Abstract – The article deals with the properties of linear combinatorial optimization problems on a set of arrangements under probabilistic uncertainty. The statement of the problem, considering the possibility of stochastic uncertainty of the initial data, is considered. The properties of the formulated stochastic problems are explored.

Keywords – optimization problem; optimization problem on arrangements; stochastic optimization.

INTRODUCTION

Actual trend of the modern theory of optimization is to study the problems of combinatorial nature ([1–14] and others). Important results have been obtained as a result of immersion of combinatorial sets in Euclidean space and study the properties of such problems ([6–14] and others). This paper is a continuation and development of a research within the Euclidean combinatorial optimization, it considers such an important class of Euclidean combinatorial optimization problems as arrangement problems.

We also should note that the problems with uncertainty, including probabilistic, attract the attention of researchers recently ([11–20] and others). Such problems arise and in combinatorial optimization. Stochastic combinatorial models can be used to describe and solve many other important practical problems. Earlier the authors [14] proposed an approach for stochastic optimization problems formalization through the introduction of the order relation on the set of random variables. Some properties of the problems on arrangements in this type of statement are discussed in this article.

I. STATEMENT OF THE PROBLEM

Let some of the initial data in the optimization problem be discrete random variables. The last will be denoted by Latin capital letters ($A, B, \ldots$), their possible values – by small ($a^i, b^i, \ldots$). In this paper, we consider only those discrete random variables among the possible values of which there is the least. We also believe that the possible values of the random variable are in ascending order, and the least value has an index 1.

Let $P(\cdot)$ denote the probability of a random event, $M(A)$ and $D(A)$ - respectively the mathematical expectation and dispersion A. Let characteristic vector of random variable A be $H(A) = (M(A); -D(a))$, $<_l$ – symbol of lexicographic order. Let us suppose also that the order on the set of discrete random variables is introduced by the following definition.

Definition 1. We will call two discrete random variables A, B organized in ascending (A preceded B) order $\prec$ (and denote this fact $A \prec B$), if one of these conditions:

1. $H(A) <_l H(B)$;
2. $H(A) = H(B)$ and there is such t that $a^i = b^i$, $P(A = a^i) = P(B = b^i)$ for all $1 \le i < t$, and:
   2.1. or $a^t < b^t$,
   2.2. or $a^t = b^t$ and $P(A = a^i) > P(B = b^i)$

is true.

We will call two discrete random variables A, B ordered by the lack of growth ($A \preceq B$) if $A \prec B$ or $A = B$.

Using introduced linear order, let us order the elements of a given finite subset Ω of the set of discrete random variables: $X_1 \preceq X_2 \preceq .. \preceq X_s$. X_s is the maximum value and X_1 is the minimum value. The definition of the minimum and maximum allows setting the optimization problem for finding the extreme elements in the given conditions. Let $X = (X_1, X_2, ..., X_k)$ — random vector. Consider a linear function $L(X) = \sum_{j=1}^{k} c_j X_j$, where $c_j \in R^1$, $X_j \in \Omega \ \forall j \in J_k$ (here and after J_k defines set of k first natural numbers), the values of the function also belong to the set Ω under all $X_j \in \Omega \ \forall j \in J_k$. Then the linear optimization problem on a sphere Q can be formulated as follows: find a pair $\langle L(X^*), X^* \rangle$ such that

$$L(X^*) = \min_{X \in Q} L(X), \ X^* = \arg\min_{X \in Q} L(X). \quad (1)$$

In particular, sphere Q can be Euclidian combinatorial set. Let $\Gamma = \{G_1, G_2, ..., G_\eta\}$ — stochastic multiset, whose elements are independent discrete random variables, $E_\eta^k(\Gamma)$ — a common set of k -arrangements from the elements of the stochastic multiset Γ [6]. Let us consider the problem of the representation (1) when $Q = E_\eta^k(G)$: find a pair $\langle L(X^*), X^* \rangle$ such as

$$L(X^*) = \min_{X \in E_\eta^k(\Gamma)} L(X),\ X^* = \arg\min_{X \in E_\eta^k(\Gamma)} L(X). \quad (2)$$

We assume that the elements of multiset satisfy

$$G_1 \preceq G_2 \preceq ... \preceq G_\eta. \quad (3)$$

Also we assume that the coefficients of objective function satisfy

$$c_1 \geq ... \geq c_\alpha > 0 = c_{\alpha+1} = ... = c_{\beta-1} > c_\beta \geq ... \geq c_k. \quad (4)$$

and

$$\text{if } c_i \neq c_j \text{ then } c_i^2 \neq c_j^2 \text{ for all } i, j \in J_k. \quad (5)$$

II. Properties of solving stochastic optimization problem on arrangements

Suppose characteristic vector satisfy the condition

$$H(G_1) \leq_l H(G_2) \leq_l ... \leq_l H(G_\eta). \quad (6)$$

Let also $\Gamma^M = \{M(G_1), M(G_2), ..., M(G_\eta)\}$, multiset Γ^M have the base $(\bar{M}_1, ..., \bar{M}_s)$ and primary specification $(\bar{q}_1, ..., \bar{q}_s)$. Let's denote

$$q_1 = 1,\ q_{i+1} = q_i + \bar{q}_i = 1 + \sum_{j=1}^{i} \bar{q}_j \text{ for } i \in J_s;$$

$$\Gamma_i^M = \{G_{q_i}, ..., G_{q_{i+1}-1}\} \text{ for all } i \in J_s;$$

$$r = \min\{j \,|\, q_{j+1} > \alpha\},$$

$$t = \max\{j \,|\, q_j \leq \eta - k + \beta\},$$

$$u_i = \begin{cases} q_i, \text{ if } i \leq r, \\ k - \eta + q_i, \text{if } i > t, \\ \beta, \text{if } i = t > r; \end{cases}$$

$$v_i = \begin{cases} q_i + \bar{q}_i - 1, \text{ if } i < r, \\ k - \eta + q_i + \bar{q}_i - 1, \text{if } i \geq t, \\ \alpha, \text{if } i = r < t, \end{cases}$$

$$k_i = v_i - u_i + 1.$$

Then for all $i \in J_r \cup J_s^t$ $(X_{u_i}, ..., X_{v_i}) \in E_{\bar{q}_i}^{k_i}(\Gamma_i^M)$ and there is minimal in solution of the problem (2) such that $i \in J_r \cup J_s^t$ $(X_{u_i}, ..., X_{v_i}) \in E_{\bar{q}_i}^{k_i}(\Gamma_i^M)$.

Let's denote $T = J_r \cup J_s^t$. If $r = t$ then $T = J_s$ and $\sum_{i \in J_r \cup J_s^t} \sum_{u_i}^{v_i} c_j X_j = \sum_{j=1}^{k} c_j X_j$. If $r < t$ then $v_r = \alpha$, $u_t = \beta$, $c_{\alpha+1} = ... = c_{\beta-1} = 0$. Thus

$$\sum_{i \in T} \sum_{u_i}^{v_i} c_j X_j = \sum_{i=1}^{r} \sum_{u_i}^{v_i} c_j X_j + \sum_{i=t}^{k} \sum_{u_i}^{v_i} c_j X_j =$$

$$\sum_{j=1}^{\alpha} c_j X_j + \sum_{j=\beta}^{k} c_j X_j = L(X).$$

Let's denote

$$\tilde{X} = (X_{u_i}, ..., X_{v_i}),\ L_i(\tilde{X}) = \sum_{j=u_i}^{v_i} c_j X_j.$$

We prove the follow theorem.

Theorem 1. If for all $i \in J_r \cup J_s^t$ $\tilde{X}^* = (X_{u_i}^*, ..., X_{v_i}^*)$ is the minimal of function $L_i(\tilde{X})$ on the set of arrangements $E_{\bar{q}_i}^{k_i}(\Gamma_i^M)$ then $X^* = (X_1^*, ..., X_k^*) \in E_\eta^k(\Gamma)$ is the minimal in solution of the problem (2).

From the properties of minimal of (deterministic) linear unconstrained optimization problem [6] on arrangements it follows that

$$M(X_j) = M(G_j)\ \ \forall j \in J_{q_r},$$

$$M(X_j) = M(G_{\eta-k+j})\ \ \forall j \in J_\eta^{\eta-k+q_t}.$$

Let us consider solving of problem of optimization of function $L_i(\tilde{X})$ on the stochastic set $E_{\bar{q}_i}^{k_i}(\Gamma_i^M)$. Since mathematical expectation of all random variables from stochastic multiset Γ_i^M are equal and (6) is true, then $-D(G_{u_i}) \leq_l ... \leq_l -D(G_{v_i})$. If $i < r$ then $c_j > 0$ $\forall j \in J_{v_i}^{u_i}$. Then it implies from (4) that $c_{u_i}^2 \geq c_{u_i+1}^2 \geq ... \geq c_{v_i}^2$. Hence there is minimal of the function $L_i(\tilde{X})$ such as $M(X_j) = M(G_j)$ $\forall j \in J_{v_i}^{u_i}$. Thus $\forall i \in J_{r-1}$

$$H(X_j) = H(G_j)\ \ \forall j \in J_{v_i}^{u_i}. \quad (7)$$

If $i > t$ then $c^2_{v_i} \geq c^2_{u_i+1} \geq \ldots \geq c^2_{v_i}$. Hence there is minimal of function $L_i(\tilde{X})$ such as $D\left(X_{v_{i-j}}\right) = D\left(G_{u_{i+j}}\right) \quad \forall j \in J^0_{k_i-1}$. Taking into consideration that $M\left(G_{u_{i+j}}\right) = M\left(G_{u_i}\right) \quad \forall j \in J^0_{k_i-1}$, we obtain that $\forall i \in J^{t+1}_k$

$$H\left(X_{v_i-j}\right) = H\left(G_{u_i+j}\right) \ \forall j \in J^0_{k_i-1}. \qquad (8)$$

If $r < t$ then $v_r = \alpha$ and (7) is true for $i = r$. Thus (7) is true for all $i \in J_{\bar{r}}$ where $\bar{q} = \min\{r, t-1\}$. Similar from $u_t = \beta$ obtain that (8) is true for $i = t$. Hence (8) is true for all $i \in J^{\bar{t}}_k$, where $\bar{t} = \max\{r+1, t\}$.

Suppose now that $r = t$ and coefficients of the objective function satisfy

$$c^2_{u_r+p_1} \geq c^2_{u_r+p_2} \geq \ldots \geq c^2_{u_r+p_k}, \qquad (9)$$

where $k = v_r - u_r + 1$, $p_j \in J^0_{k-1} \quad \forall j \in J_k$. Then $\forall j \in J_k$ the correlation $H\left(X_{u_i+p_j}\right) = H\left(G_{u_i+j-1}\right)$ is true. In the case when all the coefficients of the objective function are positive there is minimal in the solution of the problem (2) such as (7) is true for all $i \in J_s$. Hence $H\left(X_j\right) = H\left(G_j\right) \ \forall j \in J_k$.

And if the condition

$$H\left(G_i\right) \neq H\left(G_j\right) \text{ if } G_i \neq G_j \ \ i, j \in J_\eta \qquad (10)$$

is true then the arrangement

$$X^*_j = G_j \ \ \forall j \in J_k \qquad (11)$$

is the minimal in the solution of the problem (2).

Note that if elements of the multiset H are satisfy (3) but (10) is not true then the arrangement (11) may not be a minimal in solution of the problem (2).

III. Conclusions

Optimization problems on arrangements, in which the elements of the multiset are independent discrete random variables, are considered. Minimum in the problem refers in accordance with the linear order introduced to the set of random variables. Some properties of the problems on arrangements in this type of statement are discussed. Subsequent studies suggest further study of the properties of the considered problems that will allow developing methods and algorithms to solve them.

IV. References:

[1] I. V. Sergienko and M.F. Kaspshitskaya, *Models and methods of solving combinatorial optimization problems by computer*. Kyiv: Naukova dumka, 1981, 288 p. (in Russian).

[2] M.Z.Zghurovskyi and A.A. Pavlov, *Decision making in network systems with limited resources*, Kyiv: Naukova dumka, 1981, 573 p. (in Russian)

[3] A. V. Panishev and D. D. Plechistyi, *Models and Methods of Optimization in the Traveling Salesman Problem*, Zhytomyr :ZHGTU,2006, 300p. (in Russian)

[4] G. P Donets and L. M Kolechkina, *Extremum Problems on Combinatorial Configurations*, Poltava: RVV PUET, 2011, 309 p. (in Ukrainian)

[5] I. V. Sergienko, L. F. Hulianytskyi, and S. I. Sirenko "Classification of applied methods of combinatorial optimization*" Cybern. Syst. Analysis*, 2009, Vol.45, Is. 5, pp. 732-741

[6] Yu.G. Stoyan, and O.O. Iemets, *Theory and methods of euclidian combinatorial optimization*. Kyiv : Instytut systemnykh doslidzhen osvity, 1993, 188 p. (in Ukrainian)

[7] Y. G. Stoyan, M. V. Zlotnik, and A. M. Chugay, "Solving an optimization packing problem of circles and non-convex polygons with rotations into a multiply connected region," *J. Oper. Research Soc.*,2012, Vol.63,pp. 379–391.

[8] O.O. Iemets and T.M. Barbolina, *Combinatorial optimization on arrangements*, Kyiv: Naukova dumka, 2008, 159p. (in Russian).

[9] O. A. Yemets and T. N. Barbolina, "Solution of Euclidean combinatorial optimization problems by the method of construction of lexicographic equivalence," Cybern. *Syst. Analysis*, 2004, Vol. 40, No. 5, pp. 726–734

[10] O. A. Yemets and O. A. Chernenko, *Optimization of Linear-Fractional Functions on Arrangements*, Kyiv: Naukova Dumka, 2011,154p. (in Russian,)

[11] O.A. Yemets, A.A. Roskladka " Combinatorial optimization under uncertainty" *Cybern. Syst. Analysis*, 2008, Vol. 44, Is. 5, pp. 655-663.

[12] I. V. Sergienko O.O. Iemets, O.O. Yemets "Optimization problems with interval uncertainty: Branch and bound method" *Cybern. Syst. Analysis*, 2013, Vol. 49, Is. 5, pp. 673-683.

[13] O.O. Iemets and O.O. Yemets, *Solving combinatorial optimization problems on fuzzy sets*, Poltava : PUET, 2011, 239p. (in Ukrainian)

[14] O.O. Iemets and T.M. Barbolina, "About optimization problems with probabilistic uncertainty", *Reports of the National Academy of Sciences of Ukraine*, 2014, No.11, pp. 40-45. (in Russian)

[15] Yu.M. Ermol'ev and A.I. Yastremskii *Stochastic models and methods in economic planning*, Moskow : Nauka, 1979, 256 p. (in Russian).

[16] Yu. S. Kan and A. I. Kibzun, *Stochastic Programming Problems with Probabilistic Criteria*, Moscow: Fizmatlit, 2009, 375 p.(in Russian)

[17] A. B. Naumov and S. V. Ivanov, "Analysis of the stochastic linear programming problem with quantile criterion," *Avtom. Telemekh.*, 2011, No. 2, 142–158 . .(in Russian)

[18] K. Marti, *Stochastic Optimization Methods*, Springer-Verlag, Berlin–Heidelberg, 2008, 340 p.

[19] I.V. Sergienko and M.V. Mikhalevich. "Application of stochastic optimization methods to analysis of the processes of economic transformation" *System research and information technologies*, 2004, No. 4, pp.7-29. (in Ukrainian)

[20] T. Yu. Ermolieva, Yu. M. Ermoliev, P. Havlik, A. Mosnier, D. Leclere, F. Kraksner, N. Khabarov, and M. Obersteiner, "Systems analysis of robust strategic decisions to plan secure food, energy, and water provision based on the stochastic GLOBIOM model," *Cybern. Syst. Analysis*, 2015, Vol 51, No. 1, pp.125–133 .

Moving average method in epidemic process forecasting

P. E. Piletskiy, D. I. Chumachenko
National Aerospace University "Kharkiv Aviation Institute", Ukraine, pavel1997pileckii@gmail.com

***Abstract* – In this study, the application of the moving average method to the forecasting of the epidemic process is considered. As an example, the method is used to predict the incidence of ixodes tick borreliosis. To automate the calculation of the forecast, a software package was developed. The accuracy of the results was verified on real statistical data on morbidity in the Kharkiv region.**

***Keywords* – moving average methods, forecasting, epidemic process, ixodes tick borreliosis.**

I. INTRODUCTION

Among the most widespread infectious diseases in the world, a group of natural focal is distinguished, the pathogens or carriers of which are directly related to their habitat [1]. One of the most common diseases of this group in Ukraine is ixodes tick borreliosis (ITB), a group of infectious transmissible natural focal infections transmitted by ixodes ticks. Clinically, the disease occurs with a predominant lesion of the skin, nervous system, musculoskeletal system, heart muscle and is characterized by a tendency to chronic as well as latent flow. In the United States, a specialized center for the control and prevention of diseases (CDC) in Atlanta annually registers more than 5,000 cases. In European countries, according to German scientists, the number of cases can be up to 8-10 thousand per year. Intensive morbidity rates for borreliosis in France are 39.4 per 100 thousand people, in Bulgaria – 36.6 [2].

The aim of given paper is to calculate the forecast of incidence of ITB.

II. SIMULATION OF EPIDEMIC PROCESS

Simulation should be considered as the most important tool for understanding the epidemic process. The wide use of the terms "model", "mathematical model" and models themselves in epidemiology marks a step forward in comparison with the classical methodological guidelines. Simulation of the epidemic process has its own history, with its main pages are linked with mathematical models. Mathematical modeling in epidemiology is a formal description of the main elements of the mechanism of the epidemic process with the help of a system of relationships, formulas, functions, equations, etc [3]. Depending on how deeply described in the mathematical terms elements (factors, indicators) characterize the epidemic process, several classes of modeling are distinguished:

- a formal approximation, consisting in transferring knowledge of the mathematical description of outwardly similar phenomena from other areas (for example, wave oscillations) to the epidemic process;
- formal extrapolation (mainly incidence curves), which gives satisfactory results only if the factors forming the epidemic process under consideration are approximately constant;
- substantial modeling of an epidemic process with discrete or continuous flow.

Each of these modeling classes operates with its own specific set of mathematical tools that have certain limitations and indications for use. At the same time, models belonging to the same class provide a certain level of impact in the study of the epidemic process. Consequently, there is a close relationship between the task and the modeling method.

In epidemiology, modeling is used for research purposes, to predict the nature of the epidemic process and to determine the strategy for health services.

In the construction of the epidemiological model, there are several stages:

- establishment of the structure of the model based on the collected actual data on the parameters of the epidemic process (susceptibility, stability, incubation period, duration of the disease, bacteriocarrier, duration of immunity, etc.);
- mathematical formulation of the model;
- simulation on the computer of a number of variants of the epidemic process with the inclusion of various conditions that affect the spread of infection, with the purpose of choosing the optimal one.

The advantage of the modeling method is that it is possible to accurately track the sources and routes of transmission of intestinal pathogens acting in these conditions. Everything remains as in a natural epidemic process.

III. STATISTICAL METHODS OF FORECASTING

Statistical methods of forecasting are a scientific and educational discipline, the main tasks of which are the development, study and application of modern mathematical and statistical prediction methods on the basis of objective data; development of the theory and practice of probabilistic statistical modeling of expert methods of forecasting; methods of forecasting in a risk

environment and combined forecasting methods using jointly economical-mathematical and econometric (both mathematical-statistical and expert) models. The scientific base of statistical prediction methods is applied statistics and decision theory.

The simplest methods for reconstructing the dependencies used for prediction come from a given time series, that is, a function defined at a finite number of points on the time axis. The time series is often considered within the framework of a certain probability model, other factors (independent variables) are introduced, in addition to time, for example, the volume of the money supply. The time series can be multidimensional. The main tasks to be solved are interpolation and extrapolation. The method of least squares in the simplest case (a linear function of one factor) was developed by K. Gauss in 1794-1795. Preliminary transformations of variables may be useful, for example, logarithm. The method of least squares is most often used for several factors. The method of least modules, splines, and other extrapolation methods are used less often, although their statistical properties are often better.

Estimation of the accuracy of the forecast (in particular, using confidence intervals) is a necessary part of the prediction procedure. Usually, probability-statistical models for restoring dependence are used, for example, they construct the best forecast using the maximum likelihood method. Parametric (usually based on the model of normal errors) and non-parametric estimations of forecast accuracy and confidence limits for it (based on the central limit theorem of probability theory) are developed. Heuristics are also used, not based on probability-statistical theory, for example, the method of moving averages.

Modern statistical methods of forecasting also include models of exponential smoothing, moving average autoregressions, systems of econometric equations based on both parametric and nonparametric approaches [4].

In this paper, the forecasting process is implemented using the moving average method. This method is used to smooth and predict time series.

IV. Statistical Methods of Forecasting

Moving average method allows to reveal tendencies of change of actual values of parameter Y in time and to predict future values of Y [5]. The received model can be effectively used in cases if for values of predicted parameter the established tendency in dynamics is observed. This method is not so effective in cases when such a tendency is violated, for example, in natural disasters, military actions, public unrest, with a sudden change in the parameters of the internal or external situation (inflation level, commodity prices); At the radical change in the plan of activity of the company, suffering losses.

The main idea of the moving average method is to replace the actual levels of the time series under study by their average values that cancel out the random oscillations. Thus, as a result, we obtain a smoothed series of values of the parameter under study, which allows us to more clearly identify the main trend of its change.

When developing a prediction of a population of tick-borne mites, the moving average method, based on observations over 3 (or 4) previous years, is more effective than methods based on long-term observations (over 10 years or more). This is due to the fact that as a result of the application of 3-year. The moving average each of the 3 values of the indicator (over the three years) is responsible for one-third of the forecast value. With a 10-year moving average, the values of each of the indicators of the same last three years are only for one-tenth of the forecast.

Unfortunately, there is no rule that allows us to select the optimal number m of the moving average members. However, it can be noted that the smaller m, the stronger the forecast responds to the oscillations of the time series, and vice versa, the larger m, the more the forecasting process becomes more inertial. In practice, the value of m is usually taken in the range from 2 to 10. If there are enough elements of the time series, the acceptable value for m can be determined, for example, as follows:

- set a few preliminary values m;
- smooth out the time series using each given value m;
- calculate the average prediction error.
- select the value of m, which corresponds to the smallest error.

V. Automation and Program Realization

To automate the prediction of the incidence of ITB using C # programming language, a software package has been developed that allows calculating prognosis morbidity based on existing statistical data in real time. The basic configuration of the software package includes data from 2004 to 2015. The data for the years include intensive incidence rates per 100,000 population, the population's readability for tick bites, the results of studies of ticks removed from humans, the presence of Borrelia and the proportion of ticks infected by borrelia collected on the flag, the numbers of mites.

To start calculating the forecast, it is necessary to enter the years for which data are available (at least 3 years, inclusive) and for how many years ahead it is necessary to carry out the forecast (Fig. 1). Then you need to enter the data for each year or select the available values for the period 2004 – 2016 and carry out the forecast (Fig. 2). After entering the data for one year, you need to click on the "Add data" button. After filling in all the values for each year, you must click "Forecast" to complete the forecast.

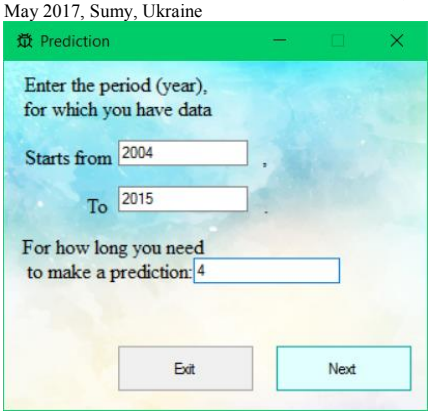


Figure 1. Initial window of the program

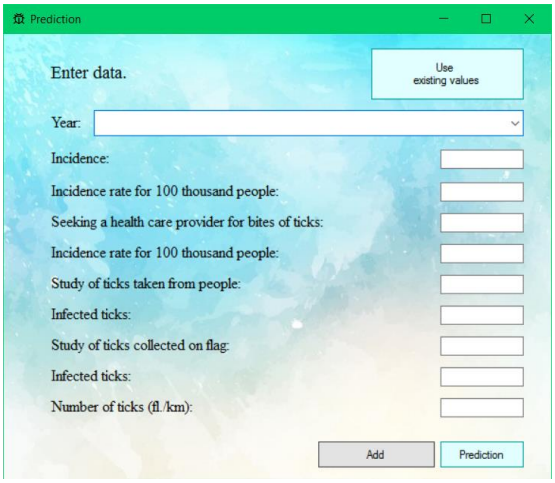


Figure 2. The data entry window for forecasting

Next, the software automatically calculates the forecast. The results are presented in the form of graphs. It is possible to look at the exact value by hovering the mouse over the point of interest.

VI. Results

Based on the calculated prognosis, it can be concluded that there is a further increase in the number of cases of ITB with a certain three-year cycle: an increase in the number of cases of ITB within 2 years with a slight decrease for the following year. On the basis of the forecast, the further increase in the number of requests for medical assistance in connection with tick bites is traced. There is no clear cyclicity. The general tendencies of the growth of incidence of Lyme disease and the population's access to medical care for bites by ticks for 100 thousand people are traced. Based on the calculated prognosis, it is possible to assume a slight decrease and stabilization of the number of ticks in natural habitats. Figure 1 shows the relationship between the level of infection by borrelia ticks in nature and ticks removed from humans. As can be seen from the figure, there is no direct connection.

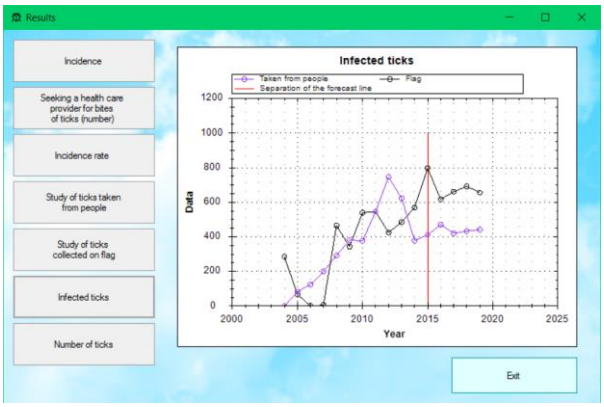


Figure 3. The level of ticks infection by Borrelia

The accuracy of calculated forecast has been verified on real statistics of morbidity by ITB in Kharkiv region.

Conclusions

In the framework of this study, an analysis of the incidence of ITB in the Kharkov region was carried out from 2000 to 2016. A software package has been developed that allows calculating the predicted incidence of ITB on the basis of the moving average method. The adequacy of the checked prognosis was verified on real statistics on the incidence of ITB.

Thus, the conducted research shows the continuing unstable epidemic situation with regard to the ITB, which dictates the need for a plan of preventive measures, the main goal of which is to reduce the incidence of people in the ITB. Virtual verification of the effectiveness of such events will be the next step in our study. Also a further area of the study is a quantitative assessment of the impact of external influences, such as public awareness, climate change, etc., on the incidence of ITB.

References

[1] M. Gulia-Nuss, A. B. Nuss and others, "Genomic insights into the Ixodes scapularis tick vector of Lyme disease", *Nature Communications, Vol. 7, Iss.* 10507, pp. 1–13, 2016.

[2] *ICD-10. International Statistics Classification Of Diseases and Related Health Problems, 10th Revision, Volume 2: Instruction manual, 2010 Edition*, World Health Organization, 2011, 201 p.

[3] Yu. Chernyshev, T. Chumachenko, D. Chumachenko, A. Tovstik, "System of Simulation of Epidemic Diseases Spreading", *Proceedings of East West Fuzzy Colloquium 2012 (19th Zittau Fuzzy Colloquium, September 5 – 7, 2012).* Institut für Prozesstechnik Prozessautomatisierung und Messtechnik, 2012. – pp.154 – 161.

[4] O. V. Yarovaya, D. I. Chumachenko, *Programming and numerical methods. Part 2: Numerical methods using MatLab and Mathcad: guidance manual for laboratory works*. Kharkiv: National Aerospace University «KhAI», 2017, 92 p.

[5] S. M. Harmsen, Y. H. Chang, S. J. Hattrup, "Simple Moving Average: A Method of Reporting Evolving Complication Rates", *Orthopedics, Vol. 39, Iss. 5*, 2016.

The features of construction the empirical description of the drop contour in automation calculations of the surface properties of the melts

Tatyana Levitskaya
Department of Computer Science Ph., D., Associate Professorıo Priazovsky State Technical University,
tlevitiisys@gmail.com

This paper considers the automation of the process of calculation the density and the surface tension of the melts according to the method of a recumbent drop. To solve the assigned task, it has been derived the empirical formulas of the analytical description of numeral solution Laplace's differential equation for the contour of a drop. It has made possible to automate fully the calculation of thermodynamic characteristics.

***Keywords*: Digital image; algorithm; Laplace's differential equation; density; least square method; surface tension; method of a recumbent drop; Delphi.**

I. INTRODUCTION

There are different methods of the experimental definition the surface tension of liquids: method of capillary rise, ring or slab detachment method, method of a recumbent drop, method of a hanging drop, drop weight method, method of the maximum pressure in gas bubble. Method of a recumbent drop gives the most exact result and nowadays it is widely used in the high-temperature searches [1]. It is typical to change the object of research during the experiment because of the interaction with structural materials of the measurement cell and furnace atmosphere for labor intensive and expensive physical and chemical measurements [2]. Therefore, it is very important to reduce the duration of the experiment, to measure all the more features and to automate experimenter labor. Such task becomes possible because of the digital equipment appearance, which can register measuring information and processes to their handling.

II. FORMULATION OF THE PROBLEM

In this method the metal drop melts on the horizontal refractory substrate or forcibly formed under the sharp edge of the crucible. Drop are photographed at a temperature of shaping and then it is measured to know its maximum diameter (*2r*) and inches under it (*h*). It is known [3] that the contour of a drop is defined by the Laplace's equation of a capillary and it has to be written like the differential second-order equation:

$$\frac{y''}{\left[1+(y')^2\right]^{\frac{3}{2}}}+\frac{y'}{x\left[1+(y')^2\right]^{\frac{1}{2}}}=\frac{(y+y_0)(\rho_1-\rho_2)\cdot g}{\sigma} \qquad (1)$$

where

σ – surface tension, N/m^2;

ρ_2 –density of the medium, where the drop is, kg/m^3;

ρ_1 –density of the medium, which is form the drop, kg/m^3;

g − acceleration of gravity, m/s^2;

y – coordinate of the point on the surface along the vertical axis;

y_0 – coordinate of the point on the top along the vertical axis.

The analytic solution this differential equation doesn't have. An approximate solution can be obtained with a predetermined degree of accuracy. All known techniques described in the literature [3] are based on using of the coupling of the system parameters with certain characteristic dimensions of the experimental drop profile. According to the sheets with theoretical drop shapes, which are calculated in advance, a connection is made between the characteristic dimensions and parameters of the drop, or by the formulas approximating the tabulated values, find the parameters of the system. Calculating the capillary characteristics by using tables is inconvenient and time-consuming.

An analysis of the theoretical aspects of the recumbent drop method has shown that the previously developed formulas and tables are either difficult to apply or not applicable at all for PC calculations, which requires transformations of the Laplace's equation to a form convenient for computer mathematical processing.

III. DESCRIPTION

To find the particular solutions of the differential equation (1) that satisfies the initial conditions $y(0)=0$; $y'(0)=0$ for different values of the capillary constant, let's use the geometric meaning of the first and second derivatives.
As a result, was obtained the expression:

$$\frac{y''}{\left[1+(y')^2\right]^{\frac{3}{2}}}=\frac{1}{R} \qquad (2)$$

which determines the plane curvature of the curve and the first derivative is equal to the tangent of the angle of inclination of the tangent to the axis *0x*. Substituting these values into equation (1), was introduced the substitution:

$$A=\frac{\sigma}{(\rho_1-\rho_2)\cdot g} \qquad (3)$$

So it was observed:

$$\frac{1}{R}+\frac{tg\varphi}{x\cdot\frac{1}{cos\varphi}}=\frac{y+y_o}{A} \quad (4)$$

$$\frac{tg\varphi}{x\cdot\frac{1}{\cos\varphi}}=\frac{\sin\varphi\cdot\cos\varphi}{x\cdot\cos\varphi} \quad (5)$$

$$R=\frac{A}{(y-y_0)-A\cdot\frac{\sin\varphi}{x}} \quad (6)$$

Equation (6) defines the radius of curvature at any point of the meridian section of the drop, but contains an undefined ratio (sin φ)/x. When approaching the top of the drop of $x\to0$, $y\to0$, $\varphi\to0$, the vertex both main droplet radius R_1 and R_2 drops of curvature equal to each other and equal to R_0. From the Laplace's equation (1) for this case was observed

$$R_0=2\cdot A/y_0$$

Having $R_1\to R_2\to R_0=2\cdot A/y_0$, this was substituted into equation (2) and found the limit of the right side of the resulting expression.

$$\frac{2\cdot A}{y_0}=\lim_{x\to0}\frac{A}{(y+y_0)-A\cdot\sin\varphi/x}=\frac{A}{y_0-A\cdot\lim_{x\to0}\frac{\sin\varphi}{x}} \quad (7)$$

$$\lim_{x\to0}\frac{\sin\varphi}{x}=\frac{y_o}{2\cdot A} \quad (8)$$

Were chosen on the meridional section line three sufficiently close to each other points M_{i-1}, M_i and M_{i+1} and drawn through them the normal lines (Fig. 1). The normal line passing through the point M_{1-1} intersects the normal line passing through the point M_i at the point O_{i-1}, and the normal lines passing through the points M_1M_{i+1} intersect at the point O_i. It was denoted the angles formed by the normal lines $M_{1-1}O_{i-1}$, M_iO_i and $M_{i+1}O_i$, respectively, φ_{i-1}, φ, φ_{i+1}.To numerically integrate the differential equation (1) and calculate the droplet shape, it was assumed that for a sufficiently small change in the current angle φ in the range from φ_{i-1} to φ_i, the radius of curvature R_{i-1} (the segment $M_{i-1}O_{i-1}$ or the segment M_i0_{i-1}) of the initial angle φ_{i-1} of the original abscissa x_{i-1} and the initial ordinate V_{I-1} does not change. In the next interval of variation of the angle φ from φ_i to φ_{i+1}, the radius of curvature R_i (of the segment M_iO_i or $M_{i+1}O_i$) or the other is calculated for a new value of the angle φi, the new abscissa xi and the new ordinate yi.

From the rectangular triangle O_i M_{i+1} P_{i+1} and O_iM_i P_i was found:

$$\Delta y_i=\left|P_i-P_{i+1}\right|=R_i\cdot\cos\varphi_i-R_i\cdot\cos\varphi_{i+1}=R_i\left[\cos\varphi_i-\cos\varphi_{i+1}\right]$$

$$\Delta x_i=\left|M_{i+1}P_{i+1}\right|-\left|M_iP_i\right|=R_i\cdot\sin\varphi_{i+1}-R_i\cdot\sin\varphi_i=R_i\left[\sin\varphi_{i+1}-\sin\varphi_i\right]$$

$$y_{i+1}=y_i+\Delta y_i$$

$$x_{i+1}=x_i+\Delta x_i$$

The graphs of the dependences of the coordinates x = x (φ), y = y (φ) of the meridional section of the drop from the current angle φ, and also the volume of the part of the drop between its vertex and the plane $y = y_0$ = const have functions

$$x=a_x\cdot\varphi^{b_x}\cdot\exp(c_x\cdot\varphi) \quad (9)$$

$$y=a_y\cdot\varphi^{b_y}\cdot\exp(c_y\cdot\varphi) \quad (10)$$

$$V=a_v\cdot\varphi^{b_v}\cdot\exp(c_v\cdot\varphi) \quad (11)$$

at the corresponding values of the parameters a_x,b_x, $c_x$$a_y$, b_y, c_y, a_v, b_v,c_v.

Therefore, for the empirical description of the numerical solution of the differential equation of the drop form (1), these dependences [4]. The coefficients of the obtained equations are found by the method of rectifying the obtained graphs and introducing new variables. Was found the ratio of the next to the previous value of any of the coordinates

$$\frac{x_{i+1}}{x_i}=\frac{a_x\varphi_{i+1}^{b_x}\exp(c_x\varphi_{i+1})}{a_x\varphi_i^{b_x}\exp(c_x\varphi_i)}=\left[\frac{\varphi_{i+1}}{\varphi_i}\right]^{b_x}e^{c_x[\varphi_{i+1}-\varphi_i]} \quad (12)$$

Take the natural logarithms from the left and right sides of this relation:

$$\ln(x_{i+1}/x_i)=b_x\ln(\varphi_{i+1}/\varphi_i)+c_x\left[\varphi_{i+1}-\varphi_i\right]=b_x\ln(\varphi_{i+1}/\varphi_i)+c_x\Delta\varphi$$

Denote $\ln(x_{i+1}/x_i)$ by $\tilde{x}$, $\ln(\varphi_{i+1}/\varphi_i)$ trough $\tilde{\varphi}$, then was obtained a linear relationship between the variables $\tilde{x}$ and $\tilde{\varphi}$

$$\tilde{x}=b_x\cdot\tilde{\varphi}+c_x\cdot\Delta\varphi \quad (13).$$

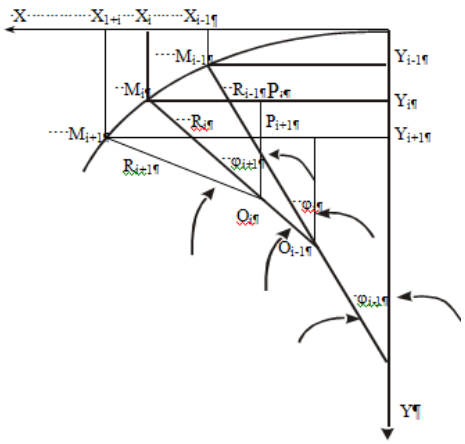


Figure 1. The scheme for constructing the drop elements for numerical integration.

While investigation unknown dependencies it is possible random errors associated with the measurement process. To reduce the effect of random measurement errors, was applied the least squares method, which allows to determine the parameters of the chosen dependence, in which the deviation from the experimental data (in this case, the calculated data) is minimal. There are

calculation results (x_1, φ_1), $(x_2, \varphi_2), \ldots, (x_n, \varphi_n)$ and the form of the function $\tilde{x} = b_x \cdot \tilde{\varphi} + c_x \cdot \Delta\varphi$. It is necessary to choose b_x *and* c_x so that the sum of the squares of the differences between the empirical and calculated values (deviations) is minimal:

$$\Phi(b_x, c_x) = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} [x_i - \tilde{x}]^2 = \min \qquad (14)$$

When the expression (7) is substituted into condition (8) it was obtained:

$$\Phi(b_x, c_x) = \sum_{i=1}^{n} [b_x \cdot \tilde{\varphi} + c_x \cdot \Delta\varphi - x_i]^2 = \min$$

To find the values b_x *and* c_x that convert the left-hand side of the resulting expression to a minimum, it is necessary to equate the derivatives with respect to zero. A function can have an extremum (min) if all its partial derivatives are zero or nonexistent.

$$\frac{\partial \Phi(b_x, c_x)}{\partial b_x} = 0 \qquad \frac{\partial \Phi(b_x, c_x)}{\partial c_x} = 0$$

$$\begin{cases} 2\sum_{i=1}^{n} [b_x \cdot \tilde{\varphi} + c_x \cdot \Delta\varphi - x_i] \cdot \tilde{\varphi} = b_x \cdot \sum_{i=1}^{n} \tilde{\varphi}^2 + c_x \cdot \sum_{i=1}^{n} \tilde{\varphi} \cdot \Delta\varphi - \sum_{i=1}^{n} x_i \cdot \tilde{\varphi} = 0 \\ 2\sum_{i=1}^{n} [b_x \cdot \tilde{\varphi} + c_x \cdot \Delta\varphi - x_i] \cdot \Delta\varphi = b_x \cdot \sum_{i=1}^{n} \tilde{\varphi} \cdot \Delta\varphi + c_x \cdot \sum_{i=1}^{n} \Delta\varphi^2 - \sum_{i=1}^{n} x_i \cdot \Delta\tilde{\varphi} = 0 \end{cases}$$

$$\begin{cases} \underline{b_x} \cdot \sum_{i=1}^{n} \tilde{\varphi}^2 + \underline{c_x} \cdot \Delta\varphi \cdot \sum_{i=1}^{n} \tilde{\varphi} = \sum_{i=1}^{n} x_i \cdot \tilde{\varphi} \\ \underline{b_x} \cdot \Delta\varphi \cdot \sum_{i=1}^{n} \tilde{\varphi} + \underline{c_x} \cdot \Delta\varphi^2 \sum_{i=1}^{n} 1 = \Delta\tilde{\varphi} \cdot \sum_{i=1}^{n} x_i \end{cases}$$

This is the final form of the normal least-squares method [5]. It was solved the system and found empirical values according to Cramer's formulas. $c_x = \frac{\Delta c_x}{\Delta}$ and

$$b_x = \frac{\Delta b_x}{\Delta}$$

$$b_X = \frac{\Delta b_X}{\Delta} = \frac{\begin{vmatrix} \sum_{i=1}^{n} x_i \cdot \tilde{\varphi} & \Delta\varphi \sum_{i=1}^{n} \varphi \\ \Delta\varphi \cdot \sum_{i=1}^{n} x & \Delta\varphi^2 \cdot n \end{vmatrix}}{\begin{vmatrix} \sum_{i=1}^{n} \varphi^2 & \Delta\varphi \sum_{i=1}^{n} \varphi \\ \Delta\varphi \cdot \sum_{i=1}^{n} \varphi & \Delta\varphi^2 \cdot n \end{vmatrix}} = \frac{\Delta\varphi^2 \begin{vmatrix} \sum_{i=1}^{n} x_i \cdot \tilde{\varphi} & \sum_{i=1}^{n} \varphi \\ \sum_{i=1}^{n} x_i & n \end{vmatrix}}{\Delta\varphi^2 \begin{vmatrix} \sum_{i=1}^{n} \varphi^2 & \sum_{i=1}^{n} \varphi \\ \sum_{i=1}^{n} \varphi & n \end{vmatrix}} = \frac{\begin{vmatrix} \sum_{i=1}^{n} x_i \cdot \tilde{\varphi} & \sum_{i=1}^{n} \varphi \\ \sum_{i=1}^{n} x_i & n \end{vmatrix}}{\begin{vmatrix} \sum_{i=1}^{n} \varphi^2 & \sum_{i=1}^{n} \varphi \\ \sum_{i=1}^{n} \varphi & n \end{vmatrix}}$$

$$c_X = \frac{\Delta c_X}{\Delta} = \frac{\begin{vmatrix} \sum \varphi^2 & \sum_{i=1}^{n} x_i \cdot \tilde{\varphi} \\ \Delta\varphi \cdot \sum_{i=1}^{n} \varphi & \Delta\varphi \cdot \sum_{i=1}^{n} x_i \end{vmatrix}}{\begin{vmatrix} \sum \varphi^2 & \Delta\varphi \sum_{i=1}^{n} \varphi \\ \Delta\varphi \cdot \sum_{i=1}^{n} \varphi & \Delta\varphi^2 \cdot n \end{vmatrix}} = \frac{1}{\Delta\varphi} \frac{\begin{vmatrix} \sum \varphi^2 & \sum_{i=1}^{n} x_i \cdot \tilde{\varphi} \\ \sum_{i=1}^{n} \varphi & \sum_{i=1}^{n} x_i \end{vmatrix}}{\begin{vmatrix} \sum \varphi^2 & \sum_{i=1}^{n} \varphi \\ \sum_{i=1}^{n} \varphi & n \end{vmatrix}}$$

After determination the coefficients c_x, b_x was found $a_x = \frac{1}{n} \sum_{i=1}^{n} \left[\frac{x_i}{\varphi_i^{b_x} \exp(c_x \varphi_i)} \right]$ that is, determined all the coefficients of the empirical dependence (3).

Similarly, we find the coefficients $a_y, b_y, c_y, a_v, b_v, c_v$ of dependences (4) and (5), which allows us to determine the volume.

The analytical description of the numerical solution of the differential equation (1) by the empirical formulas (9-11) can be considered quite accurate [6-7]. Thus, on the basis of the empirical dependences obtained, prototypes of the drop contours are obtained. This makes it possible to proceed to the realization of the next stage of identification of drop contours during the experiment and determination of the surface properties of the melts.

IV. CONCLUSIONS

An analytical description of the numerical solution of the Laplace differential equation by empirical formulas was made by using the geometric meaning of the 1-st and 2-nd derivatives. As a result of its application, a new technique for calculating the density and surface tension of melts in the sessile drop method has been developed, which made it possible to perform full automation of calculations on a PC.

REFERENCES

[1] Carla M. An automated apparatus for interfacial tension measurements by the sessile drop/ M. Carla, R. Bordi S. Ceechini // Rev. Sci. Instrum. -1991. Vol.62, №4.-P. 1088-1092.

[2] MMoser Z. Surface tension measurements of the Bi-Sn and Sn-Bi-Ag liquid alloys / Z. Moser, W. Gasior, J. Pstrum //J. Electron. Mater. -2001.-Vol. 30,№9,- P. 1109-1111.

[3] PPhysicochemical methods of metallurgical processes research / S. Philiphov, P.Arsentiev, V. Yakovlev, M. Krashennikov – M.: Metallurgy, 1986. – P. 550 (in Russian)

[4] LLvovskiy V. Statistical methods for the empirical formulas constructing / V. Lvovskiy -M.: High school, 1988. – P. 239. (in Russian)

[5] CCharles L. Lawson Solving Least Squares Problems / Charles L. Lawson, Richard J. Hanson – 1986.

[6] NN. Bakhvalov Numerical methods / N. Bakhvalov, N. Zhidkov, G.Kobelkov– M.: Laboratoriya bazovykh znaniy, 2002.–P. 632

[7] RRotenberg Y. Determination of surface tension and contact angle from the shapes of axisymmetric fluid interfaces / Rotenberg Y., Boruvka L., Neumann A. W. // J. Colloid and Interface Sci.-1993.- N 1.- P. 169-1

About the Simplex Form of the Polyhedron of Arrangements

O.O. Iemets, O.O. Yemets`, I.M. Polyakov
Poltava University of Economics and Trade, yemetsli@ukr.net

***Abstract* – The simplex form of the general polyhedron of arrangements, which is used in linear programming problems in combinatorial cutting methods is obtained and it increases the efficiency of cutting methods.**

***Keywords* – Euclidean combinatorial optimization, arrangements, the simplex form of a polyhedron, the polyhedron of arrangements.**

I. INTRODUCTION

When using linear programming problems in Euclidean combinatorial optimization as auxiliary in cutting methods [1-5], the simplex form of the polyhedron is required.

II. MAIN PART

In the report the obtaining of the simplex form for the polyhedron of k-arrangements from the elements of the multiset $G=\{g_1,...,g_\eta\}$, which is given by the system

$$\begin{cases} \sum\limits_{i\in\omega} x_i \ge \sum\limits_{i=1}^{|\omega|} g_i \quad \forall\omega\subset J_k; \\ \sum\limits_{i\in\Omega} x_i \le \sum\limits_{i=1}^{|\Omega|} g_{\eta-i+1} \quad \forall\Omega\subset J_k; \end{cases}$$

under the condition

$$g_1 \le g_2 \le ... \le g_\eta.$$

is considered.

Here and below $|\omega|$ denotes the number of elements in the set ω.

It is proved that it has the form of the system

$$\left(U-\sum_{i=1}^{|\omega|} g_i\right)\sum_{i\in\omega} X_i - \left(U+\sum_{i=1}^{|\omega|} g_i\right)Y_\omega -$$

$$-\sum_{i=1}^{|\omega|} g_i\left(\sum_{i\in J_k\setminus\omega} X_i + \sum_{\substack{\Omega\subset J_k \\ \Omega\ne\omega}} Y_\Omega + \sum_{\forall\Omega\subset J_k} Z_\Omega + V\right) -$$

$$-\alpha_{|\omega|}W_\omega^\alpha = 0,\ \forall\omega\subset J_k,$$

$$\left(U-\sum_{i=1}^{|\Omega|} g_{\eta-i+1}\right)\left(\sum_{i\in\Omega} X_i + Z_\Omega\right) - \sum_{i=1}^{|\Omega|} g_{\eta-i+1}\times$$

$$\times\left(\sum_{i\in J_k\setminus\Omega} X_i + \sum_{\forall\omega\subset J_k} Y_\omega + \sum_{\substack{\forall\omega\subset J_k \\ \omega\ne\Omega}} Z_\omega + V\right) - \beta_{|\Omega|}W_\Omega^\beta = 0,$$

$$\forall\Omega\subset J_k,$$

$$\sum_{i=1}^{k} X_i + \sum_{\omega\subset J_k} Y_\omega + \sum_{\Omega\subset J_k} Z_\Omega + \sum_{\omega\subset J_k} W_\omega^\alpha +$$

$$+\sum_{\Omega\subset J_k} W_\Omega^\beta + V = 1,$$

$$Y_\omega \ge 0;\ Z_\omega \ge 0;\ W_\omega^\alpha \ge 0;\ W_\Omega^\beta \ge 0\ \ \forall\omega\subset J_k.$$

Parameters and variables of the system are set by the following conditions:

$$X_j = x_j\cdot U^{-1}\ \ \forall j\in J_k;\ Y_i = y_i\cdot U^{-1}\ \ \forall i\in J_r;$$

$$V = u\cdot U^{-1};$$

$$\sum_{i\in\omega} x_i - y_\omega = \sum_{i=1}^{|\omega|} g_i\ \ \forall\omega\subset J_k;$$

$$\sum_{i\in\Omega} x_i + z_\Omega = \sum_{i=1}^{|\Omega|} g_{\eta-i+1}\ \ \forall\Omega\subset J_k;$$

$$\sum_{i=1}^{k} x_i + \sum_{\omega\subset J_k} y_\omega + \sum_{\Omega\subset J_k} z_\Omega + u = U;$$

$$U=\sum_{i=1}^{k} g_{\eta-i+1}+2\sum_{j=1}^{k}\left[C_k^j\left(\sum_{i=1}^{j} g_{\eta-i+1}-\sum_{i=1}^{j} g_i\right)\right];$$

$$\alpha_{|\omega|}=(|\omega|-1)U-(2^{k+1}+k-1)\sum_{i=1}^{|\omega|} g_i;$$

$$\beta_{|\Omega|}=(|\Omega|+1)U-(2^{k+1}+k-1)\sum_{i=1}^{|\Omega|} g_{\eta-i+1}.$$

III. EXAMPLE

Let us consider the example of the simplex form of the polyhedron of arrangements. Let $k=3$ $G=\{e_1,e_2,e_2,e_3,e_3\}$, that is $\eta=5$: $n=3$, $g_1=e_1$, $g_2=g_3=e_2$, $g_4=g_5=e_3$, $e_1<e_2<e_3$. That is the polyhedron of arrangements has the form

$$x_1\ge g_1;\ x_2\ge g_1;\ x_3\ge g_1;\ x_1+x_2\ge g_1+g_2;$$

$$x_1+x_3\ge g_1+g_2;\ x_2+x_3\ge g_1+g_2;$$

$$x_1+x_2+x_3\ge g_1+g_2+g_3;\ x_1\le g_5;\ x_2\le g_5;$$

$$x_3\le g_5;\ x_1+x_2\le g_5+g_4;\ x_1+x_3\le g_5+g_4;$$

$$x_2+x_3\le g_5+g_4;\ x_1+x_2+x_3\le g_5+g_4+g_3.$$

Parameters in the simplex form are:

$$U=24e_3-7e_2-14e_1,\ \alpha_1=-18g_1;$$

$$\alpha_2=15g_5+9g_4+g_3-26g_2-32g_1;$$

$$\alpha_3=30g_5+18g_4-16g_3-34g_2-46g_1;$$

$$\beta_1=12g_5+18g_4+2g_3-16g_2-28g_1;$$

$$\beta_2=27g_5+9g_4+3g_3-24g_2-42g_1;$$

$$\beta_3=42g_5+18g_4-14g_3-32g_2-56g_1.$$

The simplex form of this polyhedron is:

$$X_1(U-g_1)-Y_1(U+g_1)-g_1(X_2+X_3+Y_2+$$
$$+Y_3+Y_{12}+Y_{13}+Y_{23}+Y_{123}+Y_{12}+Y_{13}+Y_{23}+Y+$$
$$+Z_1+Z_2+Z_3+Z_{12}+Z_{13}+Z_{23}+Z_{123}+V)-$$
$$-\alpha_1 W_1^{\alpha}=0;$$

$$X_2(U-g_1)-Y_2(U+g_1)-g_1(X_1+X_3+Y_1+$$
$$+Y_3+Y_{12}+Y_{13}+Y_{23}+Y_{123}+Z_1+Z_2+Z_3+$$
$$+Z_{12}+Z_{13}+Z_{23}+Z_{123}+V)-\alpha_1 W_2^{\alpha}=0;$$

$$X_3(U-g_1)-Y_3(U+g_1)-g_1(X_1+X_2+Y_1+$$
$$+Y_2+Y_{12}+Y_{13}+Y_{23}+Y_{123}+Z_1+Z_2+Z_3+$$
$$+Z_{12}+Z_{13}+Z_{23}+Z_{123}+V)-\alpha_1 W_3^{\alpha}=0;$$

$$(X_1+X_2)(U-(g_1+g_2))-Y_{12}(U+g_1+g_2)-$$
$$-(g_1+g_2)(X_3+Y_1+Y_2+Y_3+Y_{13}+Y_{23}+$$
$$+Y_{123}+Z_1+Z_2+Z_3+Z_{12}+Z_{13}+Z_{23}+Z_{123}+$$
$$+V)-\alpha_2 W_{12}^{\alpha}=0;$$

$$(X_1+X_3)(U-(g_1+g_2))-Y_{13}(U+g_1+g_2)-$$
$$-(g_1+g_2)(X_2+Y_1+Y_2+Y_3+Y_{12}+Y_{23}+$$
$$+Y_{123}+Z_1+Z_2+Z_3+Z_{12}+Z_{13}+Z_{23}+Z_{123}+$$
$$+V)-\alpha_2 W_{13}^{\alpha}=0;$$

$$(X_2+X_3)(U-(g_1+g_2))-Y_{23}(U+g_1+g_2)-$$
$$-(g_1+g_2)(X_1+Y_1+Y_2+Y_3+Y_{12}+Y_{13}+Y_{123}+$$
$$+Z_1+Z_2+Z_3+Z_{12}+Z_{13}+Z_{23}+Z_{123}+V)-$$
$$-\alpha_2 W_{23}^{\alpha}=0;$$

$$(X_1 + X_2 + X_3)(U - (g_1 + g_2 + g_3)) - Y_{123}(U +$$

$$+ g_1 + g_2 + g_3) - (g_1 + g_2 + g_3)(Y_1 + Y_2 + Y_3 +$$

$$+ Y_{12} + Y_{13} + Y_{23} + Z_1 + Z_2 + Z_3 + Z_{12} + Z_{13} +$$

$$+ Z_{13} + Z_{23} + Z_{123} + V) - \alpha_3 W_{123}^{\alpha} = 0;$$

$$(X_1 + Z_1)(U - g_5) - g_5(X_2 + X_3 + Y_1 + Y_2 + Y_3 +$$

$$+ Y_{12} + Y_{13} + Y_{23} + Y_{123} + Z_2 + Z_3 +$$

$$+ Z_{12} + Z_{13} + Z_{23} + Z_{123} + V) - \beta_1 W_1^{\beta} = 0;$$

$$(X_2 + Z_2)(U - g_5) - g_5(X_1 + X_3 + Y_1 + Y_2 + Y_3 +$$

$$+ Y_{12} + Y_{13} + Y_{23} + Y_{123} + Z_1 + Z_3 +$$

$$+ Z_{12} + Z_{13} + Z_{23} + Z_{123} + V) - \beta_1 W_2^{\beta} = 0;$$

$$(X_3 + Z_3)(U - g_5) - g_5(X_1 + X_2 + Y_1 + Y_2 + Y_3 +$$

$$+ Y_{12} + Y_{13} + Y_{23} + Y_{123} + Z_1 + Z_2 + Z_{12} +$$

$$+ Z_{13} + Z_{23} + Z_{123} + V) - \beta_1 W_3^{\beta} = 0;$$

$$(X_1 + X_2 + Z_{12})(U - (g_5 + g_4)) - (g_5 + g_4) \times$$

$$\times (X_3 + Y_1 + Y_2 + Y_3 + Y_{12} + Y_{13} + Y_{23} + Y_{123} + Z_1 +$$

$$+ Z_2 + Z_3 + Z_{13} + Z_{23} + Z_{123} + V) - \beta_2 W_{12}^{\beta} = 0;$$

$$(X_1 + X_3 + Z_{13})(U - (g_5 + g_4)) - (g_5 + g_4) \times$$

$$\times (X_2 + Y_1 + Y_2 + Y_3 + Y_{12} + Y_{13} + Y_{23} + Y_{123} + Z_1 +$$

$$+ Z_2 + Z_3 + Z_{12} + Z_{23} + Z_{123} + V) - \beta_2 W_{13}^{\beta} = 0;$$

$$(X_2 + X_3 + Z_{23})(U - (g_5 + g_4)) - (g_5 + g_4) \times$$

$$\times (X_1 + Y_1 + Y_2 + Y_3 + Y_{12} + Y_{13} + Y_{23} + Y_{123} + Z_1 +$$

$$+ Z_2 + Z_3 + Z_{12} + Z_{13} + Z_{123} + V) - \beta_2 W_{23}^{\beta} = 0;$$

$$(X_1 + X_2 + X_3 + Z_{123})(U - (g_5 + g_4 + g_3)) -$$

$$- (g_5 + g_4 + g_3)(Y_1 + Y_2 + Y_3 + Y_{12} + Y_{13} + Y_{23} +$$

$$+ Y_{123} + Z_1 + Z_2 + Z_3 + Z_{12} + Z_{13} + Z_{23} + V) -$$

$$- \beta_3 W_{123}^{\beta} = 0;$$

$$X_1 + X_2 + X_3 + Y_1 + Y_2 + Y_3 + Y_{12} + Y_{13} + Y_{23} +$$

$$+ Y_{123} + Z_1 + Z_2 + Z_3 + Z_{12} + Z_{13} + Z_{23} + Z_{123} +$$

$$+ V + W_1^{\alpha} + W_2^{\alpha} + W_3^{\alpha} + W_{12}^{\alpha} + W_{13}^{\alpha} + W_{23}^{\alpha} + W_{123}^{\alpha} +$$

$$+ W_1^{\beta} + W_2^{\beta} + W_3^{\beta} + W_{12}^{\beta} + W_{13}^{\beta} + W_{23}^{\beta} + W_{123}^{\beta} = 1.$$

CONCLUSIONS

In this paper the simplex form of the general polyhedron of arrangements is obtained. This form of the polyhedron of arrangements is necessary for applying of Karmarkar`s polynomial algorithm in solving auxiliary problems of linear programming in combinatorial cutting methods. The increase of the effectiveness of cutting methods is to be expected, in consequence of using this form.

REFERENCES:

[1] O.A. Yemets and Ye. M. Yemets, A modification of the method of combinatorial truncation in optimization problems over vertex-located sets, Cyber. Syst. Analysis, 45, No. 5, 785-791 (2009).

[2] O. Emets' and T. Barbolina, On the solution of problems of nonlinear conditional optimization on arrangements by the cut-off method, Ukrainian Mathematical Journal, 55, No 5, 729-738 (2003).

[3] O.A. Emets and T.N. Barbolina, solving linear optimization problems on arrangements by the truncation method, Cyber. Syst. Analysis, 39, No. 6, 889-896 (2003).

[4] T.N. Barbolina and O.A. Emets, An all-integer cutting method for linear constrained optimization problems on arrangements, Computational Mathematics and Mathematical Physics, 45, No. 5, 243-250 (2005).

[5] O.O. Iemets, Ye.M. Yemets and Yu.F. Oleksiichuk, Direct cut-off method for combinatorial optimization problems with additional constraints, Cyber. Syst. Analysis, 47, No. 6, 932-940 (2011).

Optimization of Websites Advertising Budget Distribution

Pavel Mishchenko, Serhey Shapovalov
Sumy State University, Ukraine, shap@id.sumdu.edu.ua

***Abstract* – This paper is devoted to problem of the most optimal allocation of website`s advertising budget among teaser networks. We consider the possibility of improving the efficiency of distribution by means of an automated system. The main steps of genetic algorithm and integral estimation adaptation to the multiple-criteria problem solution are shown.**

***Keywords* – Teaser Network, Genetic Algorithms, Integral Estimation, Optimal Distribution.**

I. Introduction

A significant share of commercial website budget (e.g. online stores etc.) refers to advertising and customer acquisition (Table I).

TABLE I. BUDGET DISTRIBUTION BY ADVERTISING DIRECTIONS

Direction	Annual results, million UAH	
	2015	***2016***
Search (paid issue in search engines)	880	990
Banner advertising	680	710
Mobile advertising	150	200
Digital video	325	455
Total market	**2 035**	**2 355**
Another digital	200	230
Total	**2 235**	**2 585**

Vast expansion of services provision to Internet users forces to look for new ways of overcoming their irritations associated with obsessive advertising or various referral links to follow. Teasers – the next generation of banner and contextual advertising – is one of the possible solutions. A teaser is an advertising message, which contains an attention-grabbing picture and a short description, like an intriguing title [3]. To place teaser advertisement on different platforms, websites apply to services of teaser networks (TN) – intermediaries, which are responsible for advertisements allocation and perform a quality control of targeted platforms.

II. Problem Statement and Description of the Mathematical Model

The Internet advertising market is very widespread (Table II). Thus, a choice of TN is a key mission of promotion specialists. Commonly, they place advertising on several teaser networks. Then they analyze the results of their decision for a certain period, and try to find the optimal option of the budget distribution [1]. Such method is extremely time-consuming, it involves a large number of routine and monotonous tasks, that can finally lead to analytical errors and influence of personal preferences in context of human factor. Altogether, this approach negatively affects the number of new website visitors and the total cost of advertising. Therefore, to increase the number of visitors in terms of limited budget we need another tool for analyzing and allocating the budget between the TNs.

TABLE II. TEASER NETWORKS RATING

Position	Teaser network	Year of foundation	Traffic, mln clicks / day
1	AdHub	2010	90
2	Visit Web	2009	60
3	Trafmag	2012	60
4	Recreativ	2009	50
5	TeaserNet	2008	7

The purpose of this study is to optimize the distribution of the limited advertising budget of commercial website B_{sum} by means of automated system. The effectiveness of each TN will be defined as:

$$0 \le E = \frac{k_c}{k_v} \le 1,$$

where k_c – amount of unique visitors, provided by a teaser network, k_v – amount of paid displays of website advertisement. Accordingly, the number of users attracted per day from the particular TN can be calculated by the following formula:

$$k_{di} = \frac{b_{di}}{p_i} * E_i,$$

where d – is the sequence number of the day of research, i – is the sequence number of TN, b_{di} – is a share of daily budget, allocated to certain TN, p_i – is a price of 1 displaying of advertisement inside the TN.

Hence, a model of the optimal budget distribution is formulated:

$$K_{avg} = \frac{\sum_{d=1}^{d_{\max}} \sum_{i=1}^{n} k_{di}}{d_{\max}},$$

where d – is the sequence number of the day of research, i – is the sequence number of TN, d_{max} – is a research duration.

III. SOLUTION OF THE PROBLEM

In terms of optimizing the budget distribution, it is necessary to take into account the cost of advertisement allocation in various TN and its efficiency. This problem can be considered as a multiple-criteria search problem. To solve it, genetic algorithms can be adapted – a modern and fast tool for finding solutions to multiple-criteria problems.

The first step in applying the genetic algorithm is the choice of chromosomes. It is necessary to distribute the budget among teaser networks t_1, t_2, ..., t_n, therefore the values of chromosome genes will be b_{dj} – starting from B_{sum}/d_{max} to t_i. An additional constraint is $b_{dj} > 0$.

Due to the fact that genes will not be formed classically (usually they are represented by sequence of 0 and 1), it is necessary to change mutation and crossover operations. In order to save the distribution balance, mutation will be performed 2k times (even amount) for each chromosome. The crossover process will be converted into the average value calculation between two corresponding genes of the parent chromosomes

$$(b_{c1}, b_{c2}, ..., b_{cm}) = \left(\frac{b_{b_{11}} + b_{b_{21}}}{2}, \frac{b_{b_{12}} + b_{b_{22}}}{2}, ..., \frac{b_{b_{1m}} + b_{b_{2m}}}{2} \right),$$

where $b_{c1}, b_{c2}, \ldots, b_{cm}$ – are genes of the daughter chromosome;

$b_{b11}, b_{b12}, \ldots, b_{b1m}$ and $b_{b21}, b_{b22}, \ldots, b_{b2m}$ – are genes of the parent chromosomes.

The next significant step is to assess the chromosomes fitness. It is based on calculating the fitness function of each chromosome in a given population. The higher the value of this function - the higher the quality of the chromosome [4].

One of the task`s criteria is the efficiency of the teaser network. The number of times the ad is shown k_v is determined by the share of the budget allocated to the TN, while the number of visitors k_c is estimated based on the website statistics. The effectiveness of the TN is unstable and its dependence on the share of the allocated budget is nonlinear, so the statistics of visits is measured for the last 7 days (Figure 1). Due to the condition $b_i > 0$, we have the following two-dimensional function f(x,t) for each of the teaser networks used.

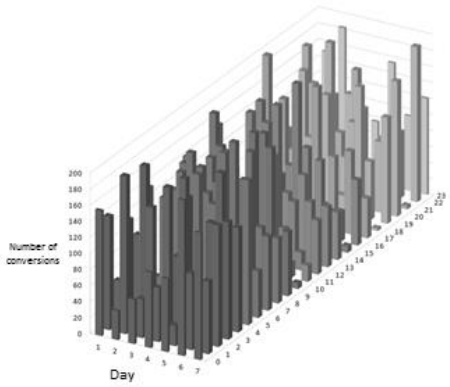


Figure 1. Statistics of conversions from the teaser network

However, the fitness assessment based on a two-dimensional random function in practice is quite complex and time-consuming. It is necessary to use the transition from a two-dimensional random function to a one-dimensional function, by definition of the integral evaluation. Such evaluation Ix(t) for the distribution density f(x,t) of a random variable x was developed in [2]:

$$I(t) = \int_{|x|>|\delta_x|}^{f(x,t)>\varepsilon_x} \frac{1}{|x|} \ln \frac{f(x,t)}{\varepsilon_x} dx,$$

where f(x,t) – the distribution density of a random variable x at time t;

ε_x – a constant parameter having the same dimension as f(x, t). In fact, ε_x determines an unobserved part of generally infinite graph of the distribution density f(x, t);

$|\delta_x|$ – a modulus of the threshold value for x, which was introduced to avoid division by zero. The values ε_x and $|\delta_x|$ are selected by the researcher depending on the problem to solve.

The use of the evaluation makes it possible to quantify the changes in the graph of random variable distribution density or the graph obtained from the histogram observations. Using the integral evaluation, we set the one-dimensional function in accordance with each teaser network. The result is shown at Figure 2.

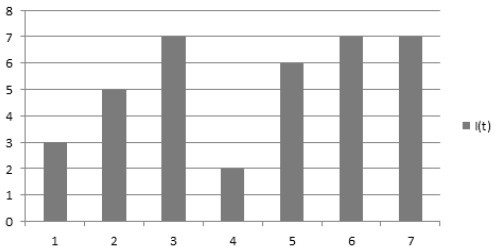


Figure 2. Integral estimation of transitions from a teaser network

As a consequence, the integral evaluation is used as one of the fitness function criteria, since it displays all information about the effectiveness of the TN, but has a simpler way of processing the results. The final fitness function has the following form:

$$\sum_{i=1}^{n} b_i * \sum_{j=1}^{7} I_{ij}(t) \to \max$$

Chromosomes with the greatest fitness function value will be the most adapted and will participate in breeding. The reason for the break would be the absence of improvement possibility in the solution of the problem. So the best solution is the chromosome with the highest fitness function value. This chromosome is the desired solution of the problem, that is, the vector of characteristics of B_{sum} optimal distribution among the teaser networks.

IV. RESULTS

The study was conducted on the 10 most popular TNs and was limited to 1 calendar month (30 days). Every day the equal part of the budget was distributed among all TNs. With the aim of comparison, the human-expert carried out manual distribution on the similar project with the same budget. The results of manual and automated distribution are available on Figure 3 and Figure 4, respectively. They represent the share in the total budget B_{sum} (in %), allocated to each of the TNs on average for the research period.

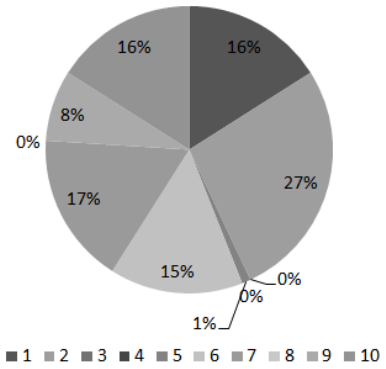


Figure 3. Manual budget allocation

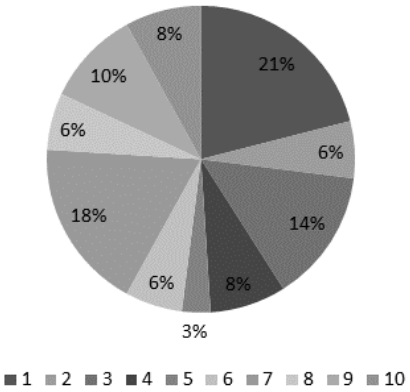


Figure 4. Automated budget allocation

The efficiency of distributions, expressed in the number of unique website visitors, is demonstrated at Figure 5. Horizontal axis represents the sequence numbers of research days (ascendingly). The vertical axis records the statistics of website visits by unique users on the relevant day.

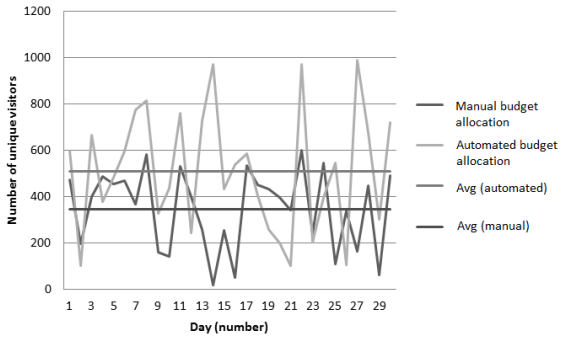


Figure 5. Statistics of site visits by unique users

The results of the automated distribution confirmed the assumption of its advantage in speed of the algorithm's reaction over the expert`s as a whole. Nevertheless, on the chart you can observe 3 days, when the results of manual distribution significantly (by 25-45%) exceeded the automated one. This is because of a problem of incomplete algorithms knowledge in comparison with the expert.

CONCLUSIONS

Optimization of budget distribution with the help of a genetic algorithm allowing to maximize the number of transitions to the advertised website in conditions of limited budget. The integral evaluation allowed simplifying and formalizing of the effectiveness estimation for each teaser network used.

The use of automated distribution increased the number of unique visitors in average by 45%, that means, it was 45% more efficient than manual. It is also possible to use automated distribution as a supplementary tool to human decisions. This approach will combine the advantages of an expert and automated methods: implement the expert's knowledge of the outside world, reduce the negative impact of the human factor and increase the speed up the reaction to changes in TN`s effectiveness.

REFERENCES

[1] Avinash Koshik, *Web 2.0 analytics in practice. Subtleties and best practices.* Dialectics Williams, 2014, p. 218. (in Russian)

[2] V.V. Avramenko, A.P. Karpenko, "Integral estimation of the distribution density of a random variable" in *Modern engineering technologies*, vol. 1. Kiev: ISMO, 1997, p. 263. (in Ukrainian)

[3] Rose Robert., *Content management. A practical guide to creating a loyal audience for your business.* Mann, 2014, p. 57. (in Russian)

[4] D. Rutkovskaya, *Neural networks, genetic algorithms and fuzzy systems.* Moscow: Hot line-Telecom, 2006, p. 452. (in Russian)

Site structure analysis and optimization

Alexander Volsky
NTUU «Igor Sikorsky Kyiv Polytechnic Institute», volsalex@yandex.ua

***Abstract* – The article contains information about current situation of site development, reasons why sites are so essential for activity of organizations and why researches of their structure improvement are so important. Short analysis of previous researches is provided, with their negative features and ways for further researches being exposed. Several ways to formulate recommendations and requirements for site structure are suggested, as well as methods for its optimization. A method for implementation of research results as software product is proposed. The information about an example of such software which performs analysis of sites of educational institution is provided too.**

***Keywords* – site structure; optimization; site type**

I. Introduction

Currently, information technologies quickly spread around the world and are used in various fields of activity. The Internet becomes an important and integral part of the life of a modern man. Now websites are the main source of information for a huge number of Internet users. Therefore, most organizations and companies create their own websites, which represent them in the Web, as well as provide information about the capabilities, features and services of the organization. Since the vast majority of sites are aimed at providing their visitors with information placed on them, and in some cases offer additional functionality, such as purchases, the questions about attracting visitors (target audience) to the site and their retention (the transfer of visitors to the site permanent users) are very important nowadays, especially if there is competition with other companies or similar information resources. This is the reason why organizations create their own website templates, define design features, content and try to organize the site in such a way that it is highly rated by search engine robots and placed among the first positions of search results in various search engines.

The user's choice of a site is influenced by several factors, among which are: uniqueness of t content, time of access to the site and search for necessary information on it, the intuitive notion of structure and functional actions, etc.

One of the most important factors is the convenience of the site structure for the user. It should be simple, logical and understandable for the visitor. If a user is forced to spend a lot of time searching for the information he needs or trying to understand organization of the page hierarchy, he is likely to simply leave the site or will not want to visit it in the future.

Therefore, research on the organization of a good and user-friendly structure of the site is essential for the effective activity of organizations at the current time.

Further in the article, the results of previous studies of the structure of the site are reviewed, their negative features and possibilities for further research and development are discussed, and ideas and methods are proposed to determine the set of requirements and recommendations for site structure that will help make it more user-friendly. An example of using research results in a software product is given as well.

II. Current situation and formulation of the problem

Based on the results of previous studies, four main types of site structure were identified:

Linear - the site does not allocate sections, and pages are linked together in sequence. A page usually contains links to previous, next and main pages. Sometimes these sites can have several page sequences, with their pages available only for certain categories of users. Linear structure is most often used in websites where online books, presentations and step-by-step descriptions of certain processes are posted.

Hierarchical - pages in the site are organized in the form of a multi-level hierarchy, which can be described as a tree structure. In the structure there is a main page that contains a menu with links to sections of the site that represent pages of next level of hierarchy. Sections may in turn contain links to subsections or relevant information. This is a universal structure that is suitable for almost any resource: home web page, organization or corporate website, information site and others.

Network - has all properties of a hierarchical structure, but also gives the ability to quickly jump from a page of one section to a page of another that is logically linked to it, or from a page belonging to one section to page of the same level of hierarchy in other section (in contrast to the hierarchical structure in which transition is possible only up and down between pages on the hierarchical levels belonging to one section). Most often this structure is used in Internet shops and information portals. However, it is also the most difficult to implement and has big amount of internal links, so it is not suitable for large sites.

Combined - combines several other structures. Often a site can have a hierarchical structure, which in certain parts of the site transforms into a linear one.

An example of such a structure is a catalogue of articles - its content can begin with hierarchically arranged pages, but the content of large sections and pages is divided into several consecutive pages.

Also, there is a defined set of recommendations for site structure organization:

Three-click rule is an unofficial rule of site structure organization. It assumes that the user should be able to find any information he needs (the web page of the site) passing through three or less links, starting from the homepage of the site.

The rule of seven - the navigation menu of the site, as well as lists of subsections of the section, should not contain more than seven items representing sections of the site.

This rule is based on the property of short-term human memory to remember up to seven entities at a time and assumes that because of this it will be more difficult for the user to search for information he needs on a site where the menu breaks this rule.

Navigation rule - for the convenience of users it is recommended to post a link to the main page of site on all its pages. It is also common to use site logo as a link to the main page.

In addition, it is recommended to create and display on each page a navigation bar that includes links to the main sections of the site (and, possibly, to the subsections of the current section).

Map rule – it is highly recommended to create a special page on a website that contains a hierarchical list of links to all its pages. This way, if a user was confused by diversity of sections and subsections, he can visit this page and search for a page which contains the information he is looking for.

However, it is a weak solution if the site structure is bad as many users are unaware that such a page exists and will simply leave the site.

This classification of types of site structures and set of recommendations are commonly used, but they are often vague and contain few specific requirements. They also don't take into account site features and goals.

For example, number of sections on certain sites is determined by number of product categories sold on it, or by number of subject areas, information about which it provides. Some portals and sites are also very big and have a large number of pages, so they have to create pages of fourth or bigger level to make navigation for their users easier.

This is why specific requirements for different kinds of websites should be made as well. The suggested ways to solve these problems are proposed in the next section.

III. RESEARCH MATERIALS AND RESULTS

At the moment, there is a set of requirements and recommendations for site structure optimization, for example, Google's Webmaster recommendations [1] or services that perform site analysis and find existing errors (for example, Netpeak Spider [2]). However, recommendations are usually given in general form and do not take into account the features and purpose of the site. Using a detailed classification of sites makes it possible to make recommendations for improving the structure more accurate and take into account the specifics of the site. However, currently, there is a problem - no single classification of sites [3,4], so introduction of a unified site classification is needed. An example of such classification is shown in Table 1.

TABLE III. MAIN SITE CLASSES

№	Site class
1	Business card website
2	Information site
3	Corporate image site
4	Online contact site
5	Online store
6	Personal project
7	Promotional site
8	Forum site
9	Online Service site
10	Web portal
11	MFA,MFS-site

Based on this classification, it is possible for each class to determine:

- recommended site structure
- the allowed number of pages
- the presence of certain sections in the main menu of the site
- special set of recommendations

To determine these parameters, a special statistical research for each class of sites should be made. The point of research is that each participant will have to find certain information on the site. To do this, he will work with two or more versions of the same site, with differences in structure, navigation, and content organization. For the study, a set of parameters is defined that serve as criteria for the convenience of the structure of each version of the site for the user. Such criteria can be:

1) Information search time – it is one of the main indicators of the quality of the site structure. Any flaws in it, such as a big nesting level of pages, confusing navigation, illogical organization of hierarchy, automatically increase this indicator. Accordingly, the lower its value, the better the structure of the site is.

2) The user's personal rating - a subjective user's estimation of the convenience of the site structure, which takes the value from 1 to 10 and is determined by the user himself after working with the site version.

Other possible criteria are given in [5].

Also, it is possible to determine the type of site structure and requirements for it based on the number of pages in it. To do this, at the initial stage, recommendations of web studios can be used, for example "Antula" [6] (Table 2) with further analysis of the type of site and user behavior when the structure is changed.

TABLE IV. RECOMMENDED SITE STRUCTURE

№	Number of pages	Recommended structure
1	<40	One-level structure
2	40-700	Two-level structure, with third nesting level possible for 10-15% of pages
3	700+	Three-level structure, with fourth nesting level possible for 10-15% of pages

Based on the results of such researches, the specific requirements for the optimal structure of the site of each class can be determined. In the future, software product implementation of research results can be made as a service that analyzes site structure of the chosen site, compares its parameters with parameters determined by the requirements for the structure of this class of sites, and generates a set of recommendations for analyzed site structure optimization.

For personal research, a class of information sites and its subclass - sites of educational institutions was selected. The research was made on the websites of the faculties and departments of the Igor Sikorsky Kyiv Polytechnic Institute. The reasons for this choice were:

1) a large number of faculties and departments in the university, so there is a big number of relevant sites;

2) significant differences in the structure and design of these sites;

3) the necessity to update and optimize these sites

The importance of the research is also increased because of the fact that the sites quality directly affects the opportunities for attracting new enrollees, the sites position in the search engines, and the position of the educational institution in the ratings that take into account its electronic resources, such as Webometrics [7].

Target audience of such sites includes enrollees and their parents, students, as well as employees and managers of the educational institution. The site must provide them with the information necessary for each category. Therefore, the main criteria for analyzing these sites are:

1) number of pages of the site and accordance of site structure to recommended;

2) nesting level of pages;

3) number and length of items in the main menu of the site;

4) link quality - the presence in the link text of words associated with the content of the page

For the research, special software "Edu Spider" was created. The structural scheme of the program is shown in Fig. 1.

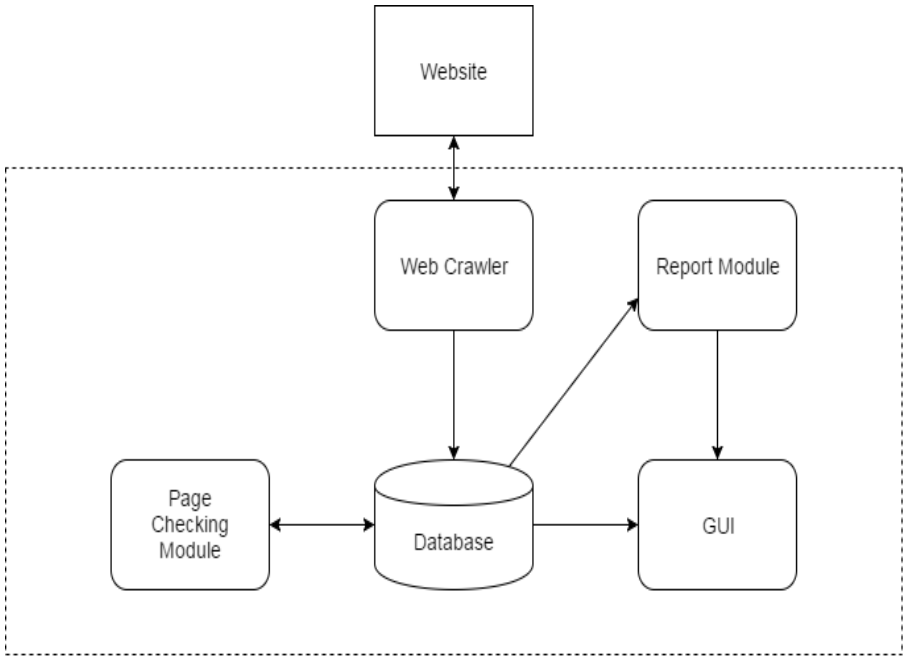


Figure 1. Structural scheme

The web crawler performs crawling of site internal pages and gathers the necessary information that is saved in the database. After that, the page checking module checks whether the site structure and certain page parameters match requirements and recommended values. If there are disparities and errors, the module adds information about them to the database. Information from the database, if necessary, is processed in the report module and displayed in the user interface, giving him information about the results of the site analysis.

Also, the program builds a tree graph of site pages, with problem pages being highlighted (Figure 2).

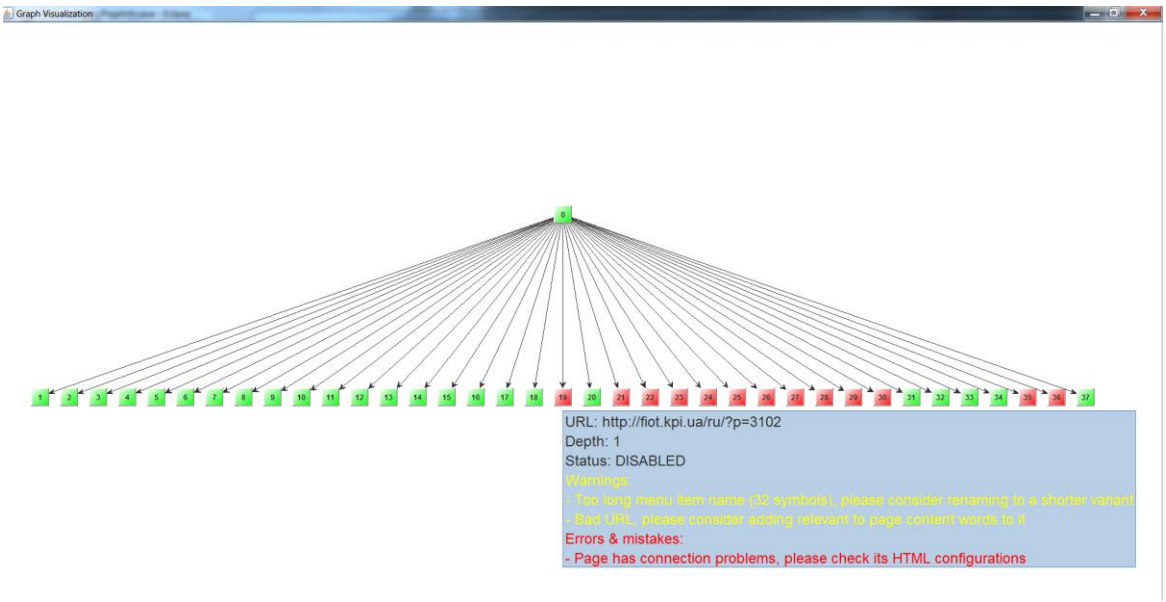


Figure 2. Tree graph of site pages

Each page of the site is represented by a rectangular-shaped node of a tree with a unique page identifier. Pages where no problems or inconsistencies were found are highlighted in green. Problem pages are highlighted in red. When hovering over a page, the user can see the general information about the page: its link, nesting level and status. Nesting level shows the minimum number of transitions from the site main page needed to get to this page. The status determines ability to connect to the page. The problem pages additionally show information about detected inconsistencies and recommendations how to remove them. Possible errors and inconsistencies of the recommended page parameters are indicated in yellow. Explicit errors and violations of requirements are indicated in red.

CONCLUSIONS

In this article, the main features and problems of site structure analysis and optimization were shown. Several methods have been proposed to solve the problems discussed. They include the introduction of a general classification of sites, the definition of specific requirements for each class of sites, as well as methods that can be used to determine the appropriate structure for the site based on the number of its pages. These solutions can make the requirements more specific and take into account the specific features and purpose of the site, which will complement the set and correct the main shortcomings of existing requirements.

Also, the directions for further research and some features of their conduct were indicated. Based on the results of these studies, it is possible to create special software that will be able to automatically analyze the site structure and offer opportunities for its optimization.

Based on the research, special software was developed that performs the analysis of the websites of the higher educational institution Igor Sikorsky Kyiv Polytechnic Institute and offers recommendations how to remove discovered shortcomings.

REFERENCES

[1] *Google Webmaster Guidelines.* (n.d). Retrieved from https://support.google.com/webmasters/answer/35769?hl=en
[2] *Netpeak Spider: Your Personal SEO crawler*. (n.d). Retrieved from https://netpeaksoftware.com/spider
[3] Koksharov, S. (2010, June 24). *Internet Marketing site classification*. Retrieved from https://www.seonews.ru/analytics/klassifikatsiya-saytov-v-internet-marketinge/ (in Russian)
[4] *Yandex Catalogue site classification*. (n.d). Retrieved from https://yandex.ru/support/catalogue/basic-info/classification.html (in Russian)
[5] Kosarekov, I. (2013, July 2). *Site quality criteria*. Retrieved from https://blog.mcsite.ua/quality-site.html (in Russian)
[6] *About Antula Web Studio.* (n.d). Retrieved from http://www.antula.ru/about-firm.htm (in Russian)
[7] *About Webometrics.* (n.d). Retrieved from http://www.webometrics.info/en/About_Us

Methods of automated content-based answer search for automation first level of technical support

Oleg Nestyuk[1], Ekaterina Lischuk[2]
1National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ukraine, nestoleh@gmail.com,
2National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ukraine, lishchuk_kpi@ukr.net

***Abstract* – Automated content-based answer search is proposed to use at the implementation of the first-line technical support for the product.**

The main task is to inform users about product and to solve the most common problems in product work. Question-answering system and intelligent chat bot are proposed as technologies that can be used in implementing this task.

***Keywords* – automated search, question answering system, chat bot, technical support**

I. THE ACTUALITY OF THE THEME THE TOPIC

The search of answers according to the contents is kind of informational search when it is possible to find the short answer on the question that is set by user in native language. Contemporary search systems for the network Internet let us to get the list of documents (pages), that contain necessary information and are regulated by relevance. Herewith the receiving of necessary information from the pages is let by the user. The systems of automated answers search, in contrast to traditional search systems, get question in native language (Ukrainian, English, etc.) but not the set of key words and return short answer, not the list of documents and references.

Consider the question of actuality of automated content-based answer search is worth to define the particularities in sphere, where such software can be used. The features of sphere of usage of automated search of finding answers are the next points:

- Limited basic of answers, the sphere where questions can be put should be limited by some frameworks (by subject, connection with some object or other);
- The possibility to formulate definitely correct answer on the question that can arise in the user of system;
- The possibility to reduce communication of user with software to creation of inquiries and reflection of answers in written form;
- The understanding of user that fact, that he communicates with automated system, but not with living being, that factor is necessary for increasing of effectiveness such kind of communication.

To all this requests correspond the part of existing system of technical support for different kinds of productions, including software. The essence of technical support of the first level goes to taking the question from the user, filtration of this question from unconnected information, summarizing to the form of one of the known question in the database and presenting of this answer to the user.

II. GENERAL SCHEME OF THE WORK

The general scheme of the work of automated content-based answer search can be presented by following steps:

Step 1. Getting the question from the users. User inserts the question in written form in the window of the chat.

Step 2. Checking is the question is not the instruction. The questions that include some words that fulfill the expressions that are defined in system as regular expressions can be interpreted as instructions. These instructions are made on the level of bot.

Step 3. The analysis of the question. The analysis of text that is introduced by user is held and is checked if introduced text can be interpreted as question.

Step 4. Reconstruction of the question. The question is reconstructed in suitable for work with questions-answering system format with taking into the consideration the contents of the conversation.

Step 5. The search of answer. Transference of the question to question-answer system and getting answers or informing about the absents of answer.

Step 6. Outputting the answer. Outputting the answer to the user. [2]

The basic descriptive scheme of work of automated content-based answer search is represented on the figure 1.[4]

III. THE SEARCH OF ANSWERS

In getting question from the user, the first global stage is search of answers on the question. The search of answer is the sphere of responsibility of question-

answering system, which is the component of general model of automated content-based answer search. For this is suggested the usage of aim searched system for searching by key words. From the question of user, putting in the native language, is necessary to pick out key words that form the essence of question. After receiving the key words it is necessary to settle the following problems: the search of answer on the question and formulation the answer for the user. For formalization of problems of both groups is offered to use the math set of discrete math, syntactical analysis and neuron's network.

The general scheme of the work of answer search is represented on the figure 2. [3] In this stage can be depicted the following stages:

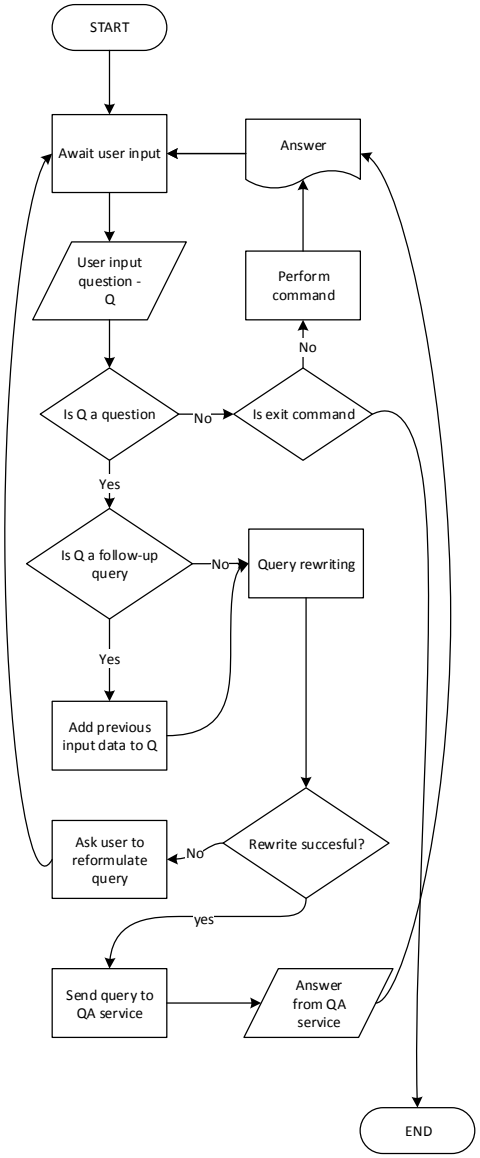


Figure 1 – Flowchart of the algorithm of automated content-based answer search

Step 1. The analysis of the question. On this stage it is necessary to settle of searching requirements for subtask of informational search. As a rule for this is used the parts of the sentence without interrogative parts. Also the search requests are formed with help of paraphrasing of interrogative to statement sentence.

Step 2. Informational search. In way of source of facts in this work is suggested to use the local database with text's fragment those are the answers on conditional questions. That's way the search of information is settled to the search of key words in indexation basis.

Step 3. Separating the answer. The stage of separating of answer don't need the process of separating the answer from the textual fragment, just as all the fragment is the candidate in the answer on the question. But for the final creation of answer it's necessary to filtrate receiving fragments and to rank the answers that are left after the filtration. [1]

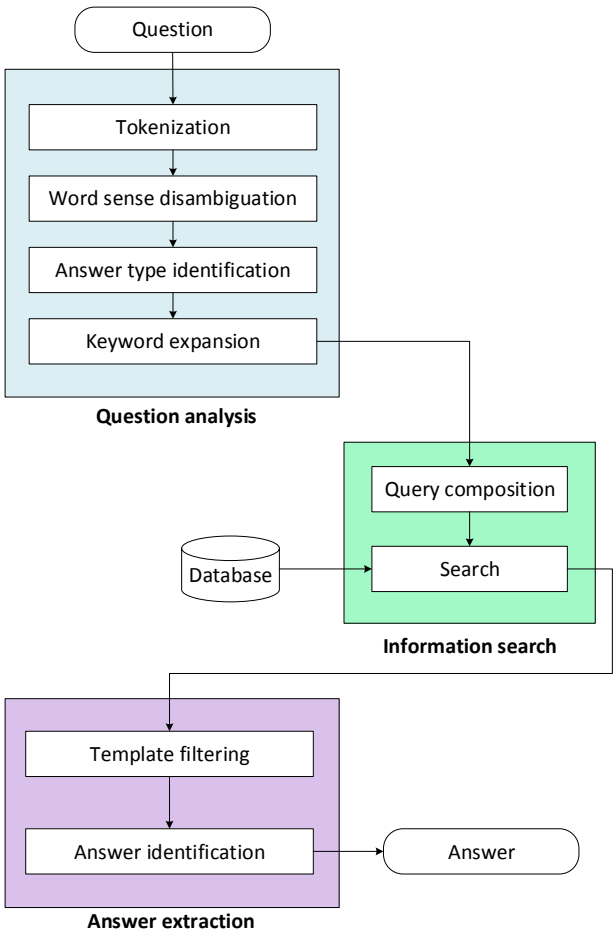


Figure 2 – The general scheme of search for automated content-based answer search

IV. PECULIARITIES OF REALIZATION OF CHAT BOT

The communication with user is executed in process of written dialogue, for this part is responsible chat bot. Virtual interlocutor or chat bot is computer program, that is created for imitation of communicative behavior of

person during the communication with one or more users. Modern interlocutor-programs are only the attempt of imitation of clever dialogue with machine. Even the most successful program of this form can't think logically as a human, make thinking operations: analyses, synthesis, comparison, classification, abstraction, summing up, concretization of information which is included in its basis of knowledge. But herewith these programs can participate a simple dialogue, that make a wide field for their potential application.

One of the best variants is realization of chat-bot with help of usage of recurrent network. But also possible is the usage of other types of chat-bot, the selection of maximum effective model will depend on sphere of usage and additional requirements for the work of chat-bot (except the supporting of simple dialogue with users). The scheme of work of chat-bot is pointed on the figure 3. The main steps, which chat-bot should do irrespective of its architecture are the following:

Step 1. Checking, as if introduced text is the command.

Step 2. Search of answer on introduced text in the part of chat-bot.

Step 3. Formation of final question and transmission of it to question-answer system. [5]

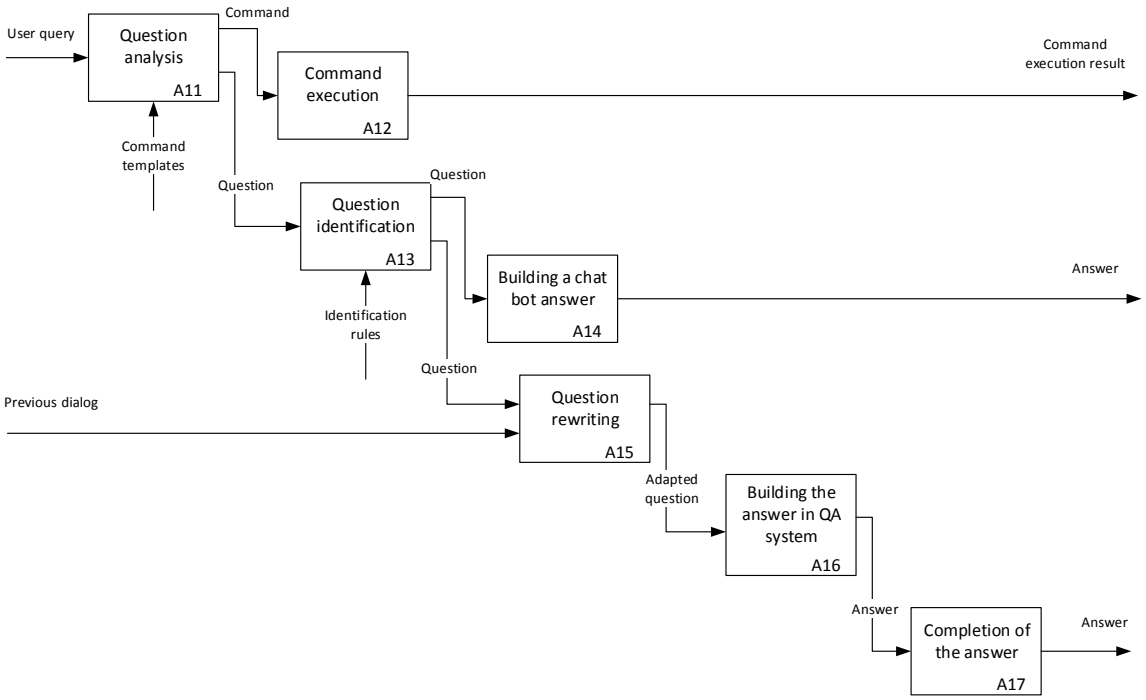


Figure 3 – Chat bot work scheme

CONCLUSIONS

Proposed method of work of automated content-based search with usage of chat-bot will let to automatize the process of work of the first level of technical support and substantially reduce expenses of company on this kind of work.

Communication with user is held in written form with means of chat-bot. Chat-bot is responsible for communication with user, and it makes main three problems: identification of commands, answering on not thematic questions and final formulation of request to question-answering system.

For search of answer is responsible question-answering system. Method is included in three stages: the analysis of the question, informational search and separating of answer.

REFERENCES

[1] А.А. Соловьёв, "Синтаксические и семантические модели и алгоритмы в задаче вопросно-ответного поиска" *Conference: Proceedings of the 13th All-Russian Scientific Conference "Digital libraries: Advanced Methods and Technologies, Digital Collections"*, RCDL 2011, Voronezh, Russia, October 19-22, 2011.

[2] Y. Liu, M. Liu, X. Wang, L. Wang, J. Li, "PAL: A Chatterbot System for Answering Domain-specific Questions," *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pp. 67–72, August 2013.

[3] B. M. Matteo; M. Negri; R. Prevete; H. Tanev, "Multilingual Question/Answering: the DIOGENE System," *10th Text Retrieval Conference proceedings*, pp. 313–321,2001

[4] J. Smith, "IQABOT : A Chatbot-Based Interactive Question-Answering System", Techical report, 2010
S. Quarteroni, S. Manandhar,"A Chatbot-based Interactive Question Answering System" *Proceedings of the 11th Workshop on the Semantics and Pragmatics of Dialogue*, pp.83-90, 30 May – 1 June 2007

Modeling business process of labor intensity calculating the machine-building equipment's production

V. Yatsenko, S. Vashchenko, Yu. Parfenenko
vashchenko.svitlana@gmail.com

***Abstract* – the main aim of this article is the solution of actual reengineering business process problem of labor intensity calculation, which is carried out at the JSC Nasosenergomash enterprise. The model of the existing process is formed in this work, which analysis allow to identify factors that reduce the effectiveness of the calculation. As a result of the research was formed a renewed model of the updated business process, without revealed faults. The technology, which will allow the reengineering business process is proposed and will lead it to the new model.**

***Keywords* – database, Intermech, information system, conversion, DBF , MSQL, software module, reengineering, labor intensity.**

I. INTRODUCTION

The activity basis of any organization is represented by its business processes, which are defined by the purposes and tasks of business entity activity. Processes provide realization of all kinds of the enterprise activity connected with production of goods or services by the business entity.

Priority in the market is one of indicators of effective work of the enterprise and its functioning in modern competitive conditions, mechanisms is necessary which can simplify difficult processes or modernize outdated.

Business process of labor intensity pumping set calculation is carried out at the Sumy pump plant JSC Nasosenergomash. Calculation of labor input of pumping set production is very important process at the enterprise. It allows to create data which calculation of salaries of the workers involved in production of one machine-building unit.

Before putting into operation of computer facilities these calculations were carried out on paper. With adoption of computer facilities, labor calculations were made by means of computer technologies, special software was developed for department of the organization of work and salaries.

With development of information technologies there was a production automation. PDM system Intermech was introduced in the enterprise that allowed to start electronic system of document circulations. However process of labor intensity pumping set calculation was almost without changes.

In the outdated database large volumes of data which are necessary for labor intensity calculation are already saved up, but the structure and a format of saving data in the outdated database differs from Intermech.

It makes impossible a direct transferring of data to one system, a common information space. For such circumstances completion of calculations cannot be automated in full, and in the course of data transferring can be a mistake.

Therefore the aim of this research is increasing business process of labor efficiency calculation of pump equipment production in the enterprise.

II. THE ANALYSIS OF EXISTING BUSINESS PROCESS

It is known that guarantee of successful reengineering realization of any business process is well-grounded analysis of the existing condition of process. So, at first it was study in detail the business process of labor intensity calculation how it is realized by employees of the enterprise for this time.

According to the IDEF0 it is made structurally functional model AS - IS, one of levels of it is given in fig. 1.

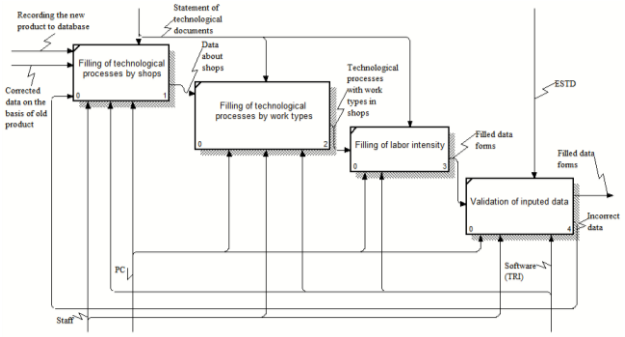


Figure 1. Model As-Is of data transfer process in the manual mode

The analysis of the existing model showed that now there are several factors which considerably reduce the given business process efficiency.

The database which keeps all information on calculations represents a set of tables in the DBF format and index files to them that are saved in different catalogs in the hard disk drive of the server.

The structure of the database is not perfect. For example, for annual release of the equipment it is created

the separate table. At the same time in such cases duplication of information is observed. It leads to the fact that the number of tables files considerably grows from year to year.

The control system of the TRI database works in MS emulation mode - DOS. So for the work of a program in some computers is installed out-of-date Windows XP, in others - the virtual machine with Windows XP. In the course of work with the program loading of the processor fluctuates in a section of 90-95%, and it complicates doing of parallel tasks.

The structure of the database imperfect, in particular it is made the new table for annual set's production.

There are no autoincremental field in table, also links between tables are absent. Labor intensity calculation for set details is not provided only on some technological processes.

An employee in the manual mode search analogs and enters additional data in the table of розцеховочний маршрут of the set which are obtained from the "Techcard" software which is a part of PDM – system Intermech.

The results of labor intensity calculation received of "TRI" are transferred to the Intermech system in the manual mode.

In general it is possible to make a conclusion that business process of labor calculation of machine-building production actually does not correspond to the updated concept enterprise's work, but it can't refuse it because the data which were saved up during a long time are used for different analyses and calculations. So it needs modernizations.

III. MODEL OF UPDATED BUSINESS PROCESS

It is necessary to make an advanced To - Be model of business process for successful reengineering realization of it and factors elimination earlier defined. The main aim of the new business process organization is the accomplishment's transfering of all its stages by means of Intermech system modules. The updated structural – functional diagram of business process is shown in fig. 2.

All the work will be within single support system of Intermech production in proper modules unlike existing business process. This is simplify the given business process' completion.

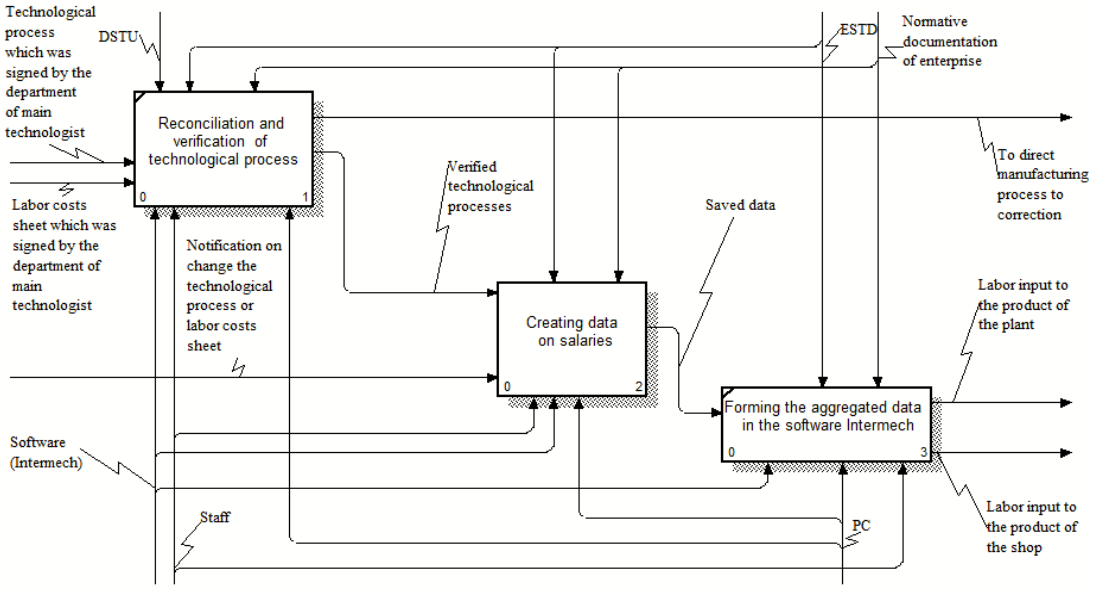


Figure 2. Model To-Be of the Business Process

For example, if an employee finds errors in documents by data verification, when he opens the previous registrations about some set, or the set very similar to specifics production, then the given documents return in system with the reason of returning.

This process allows to save time for coordination of documentation which was spent for preparation service notes, signatures and stamps gathering and also for documents transfer between departments.

All necessary reports are formed by the built means, are registered in the system and are saved in electronic archive automaticaly.

The presented process of labor intensity calculation correspond to the working concept of the enterprise in a common information space of an automated control system of Intermech.

At first it allows to refuse completely the outdated software and virtual machines that free computer resources.

Secondly – transfer to single document system will allow to see all made changes in registrations in the shortest time that in turn reduces time for data processing and preparation necessary analyses and reports.

IV. BUSINESS PROCESS' REENGINEERING

The existing business process' reengineering is offered to be executed due to realization of several stages. Generally business process' reengineering scheme which is considered is shown in fig. 3.

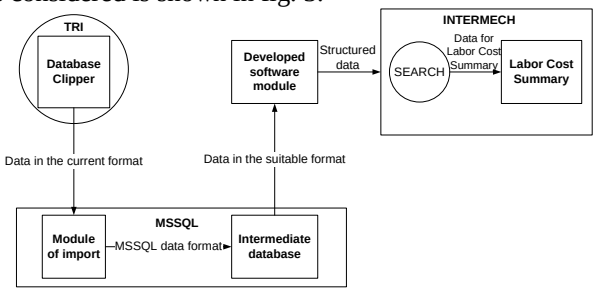


Figure 3. Scheme of the reengineering process

The aim of this research is to find differences in structures of databases and a way of reduction of an old data format in a database which is used in Intermech.

A key role in reengineering of labor intensity process calculation is played by transferring of data to a common information space as this step provides future opportunity to operate with data in the Intermech system. Reorganization of business process has to begin with it.

At the first stage it is necessary to investigate structures of these tables TRI and Intermech in details. The aim of such research is to find differences in structures of databases and a way of reduction of an old format of data in a database format which is used in Intermech. Directly, there is no instrument of import of data from DBF format in a control system of the Search database. Therefore the following scenario is proposed. At first it is necessary to create the intermediate local database in Microsoft SQL Server as all databases of the enterprise use this format. To import data from tables of an old format. Tables import occurs without the previous structuring of the existing data. However in the course of the import completion by the built means of the program SQL Server Management Studio module it is possible to edit structure of database tables, leading them in correspondance with the database which is used in Intermech. In particular, it is necessary to lead types of these fields of the existing tables to the corresponding types of these new created tables as formats of fields in databases are differ and an error of data import ensue. The corresponding adjustment are made in the manual mode in the editor of the table of the import module of MSQL. It is also necessary to unite the same tables in one. For the correct tables organization in the intermediate database it is necessary to create key fields in tables and connect registrations between tables by keys. Then it is necessary to transfer information from the intermediate database to a control system base Intermech. Considering data size which are kept in the old database and the probability of emergence of mechanical mistakes, manual transferring of data is economically and logically inexpedient. So it is expedient to develop the software which would automate the last stage of data import. The program module has to carry out such functions: revision of tables structure of the intermediate database;key fields assigment;tables linking; import all information which contains in tables of the intermediate database to the database of the Intermech system, providing avoidance of empty registrations and data duplication. After accomplishment of the described actions data will be imported to Search control system and available to use by other Intermech modules.

CONCLUSIONS

In a research is proposed the technology actual problem's solution business process' modernization of labor intensity calculation pump set production which is carried out at the Sumy pump plant JSC Nasosenergomash.The research of the existing business process condition is conducted with use of modern methods of the reengineering process organization. The structurally functional model of the studied process is constructed according to the IDEF0 standard. The well-grounded system analysis of the received model allowed to allocate those stages and factors which reduce efficiency of given works accomplishment. As a result, business process completion of the existing technology leads to considerable waste time for information transfer between the different software which accompanies this process. The work is made for modeling such business process that negative factors would be eliminated.

A result of the research is an updated business process model of labor intensity calculation of the machine-building equipment production. A new model advantage is ensuring business process accomplishment of PDM in a common information space PDM system automation production Intermech which functions at the enterprise. It is proposed a technology by authors for realization of the constructed model which consists of several separate stages and will allow to execute the business process reorganization. Business process modernization of labor intensity calculation of pump equipment production for the proposed technology will allow to reduce time and expenses on carrying out the corresponding calculations. Also it will promote simplification and acceleration of process of document flow between enterprise's departments.

REFERENCES

[1] Yatsenko V.P., Vashchenko S.M., Matvienko V.V. Analiz biznes-protsesu rozrahunku trudomistkosti vigotovlennya virobu // Sbornik nauchnyih trudov SWorld. – *vol*.2. – Ivanovo: «Nauchnyy mir», 2016. - С. 41 – 46.

[2] Yaki problemi ukrayinskogo virobnitstva mozhna rozv'yazati za dopomogoyu IT [Electronic resource] – Available at: http://forbes.net.ua/ua/opinions/1412861-yaki-problemi-ukrayinskogo-virobnictva-mozhna-rozv-yazati-za-dopomogoyu-it

[3] Kompleks sredstv avtomatizatsii tehnologicheskoy podgotovki proizvodstva [Electronic resource] – Available at: http://www.intermech.ru/techcard.htm.

Investigation of industrial robot-manipulator computer model motion control

Y. Mashyn [1], E. Kuznetsov [2], Vira Shendryk[2]
[1] Opole University of Technology, Poland
[2]Sumy State University, Computer Science Department, Ukraine
Sumy, Ukraine, kuznetsov@opm.sumdu.edu.ua

Abstract – This article provides a technology for develop a physical and mathematical simulation of the robot's computer model. The purpose of the work is to learn how to define the movement of robot elements. A complete, controlled computer model of an industrial robot manipulator has been created. Model control is implemented using the modeling environment software. The technology developed is used to master the basics of robotics.

***Keywords* – robot; trajectory; positioning; coordinates; nodal points; movable link; software.**

I. INTRODUCTION

An objective reason for the arising and development of modern Robotics is a constant need flexible automation for production, for removing a person from direct participation in native production, for decrease disadvantage of traditional means of automation.

Application of robot opens up broad prospects for developing innovative technological processes without constrains, that direct participation human in them [1]. It means human physical limitations (load, speed, accuracy and repeatability manipulation), and its working conditions. Today direct human involvement in the process is often a major obstacle to the further intensification of production and creating new technologies.

Industrial robots are widely used in various fields of economy: research, transport, agriculture, medicine, in the service sector. The introduction of Robotics gives considerable technical and economic effect of increasing the productivity, technological processes, improve their product quality. Also positive is the release of a person from a difficult, dangerous and monotonous work. Complex automation of production and flexible automated productions is one of the main directions of introduction of Robotics and largely determines the modern Robotics growth rate.

Industrial robots are very expensive, and use them for educational purposes cannot afford any educational institution. However, there is an opportunity to explore controls and dynamics of industrial robot in its computer model. This model is not demanding in maintenance, can be changed easily, and can't corrupt when receiving invalid movement commands. Using the model, you can quickly optimize the robot-prototype work for their specific application and environmental conditions.

The main objective of this study is to develop a computer model of a controlled robot-manipulator for use in educational process in training students of computer specialties.

Information technology of creation the computer model of a robot-manipulator allows to investigate the movement of work elements for any robot-prototype trajectories, and to analyze the design features of the robot in terms of optimal execution of operational tasks. This technology allows to accelerate the assimilation of the foundations of Robotics and gives the opportunity to apply the skills of programming the controllers in practice.

II. OBJECT AND ENVIRONMENT OF SIMULATION

The object for computer simulation is a robot-manipulator "FANUC LR Mate 200iC" (Fig. 1). This compact high-speed robot with high positioning accuracy is ideal as a tool for studying the control of robotic systems. Equipping the robot-prototype by special tools allows to learn the various aspects of the use of such systems in the workplace. Robot-manipulator is operated by controller R30iA [2]. It has a closed mechanical design that hides the mechanisms of drives and electronic components. A computer model created by using an Autodesk Inventor software. This software allows you to easily program the model elements moving along specified path. One of the interesting features of Autodesk Inventor is the creation of custom software components that implement the automate elements of the most frequently performed tasks [3]. These software components are implemented on the basis of application programming interface (API) and advanced object modeling. The most effective way to create custom applications based on AI is the use of Visual Basic for Application programming environment. Autodesk Inventor API determines the modus operandi of the programmer and the selection of appropriate programmatic functionality of operation environment [3]-[4]. In other words, Autodesk Inventor API allows you to create an optimal sequence of commands for Autodesk Inventor [4].

Figure 1. Robot-manipulator “FANUC LR Mate 200iC”

The study develop a robot-manipulator simulator, which includes a computer model of the robot and its movement control. In order to achieve the objectives of the study were assigned the following tasks:

1. To develop a mathematical model of prototype robot-manipulator elements movement.

2. To analyze the programming techniques of the movements of computer models by Autodesk Inventor tools.

3. To develop a motion control algorithm of a robot-manipulator.

III. PHYSICAL MODEL OF ROBOT-MANIPULATOR MOTION

Planning the motion trajectory of the robot-manipulator elements is the task of control method selecting to ensure the movement of the manipulator along the specified trajectory. Before starting the manipulator, it is important to know:

1. Are there obstacles in the way of the robot elements movement?

2. Are there constrains on the mobility of robot elements in its area of operation?

When planning trajectory, one of two approaches is used:

1. A set of conditions for positioning, speed and acceleration of the generalized coordinates of the manipulator's working tool in the specific (nodal) points of the trajectory is assigned. The trajectory scheduler chooses from a class of functions (typically among polynomials a function that will route trajectory through the cue points and satisfy the specified constraints in the trajectory. Constrains and plan the trajectory are defined in generalized or Cartesian coordinates.

2. The desired motion trajectory of the manipulator elements is assigned as an analytically described function. The planner develops an approximation of a given trajectory with a mathematical function in generalized or Cartesian coordinates.

Planning the trajectory in generalized coordinates has the following advantages:

1. Specifies the change of coordinates that are directly control the movement of the manipulator.

2. The trajectory can be planned in real time.

3. The trajectories in generalized coordinates are easier to define.

The disadvantages of this planning are difficult to determine the position of the manipulator elements during the movement process. This is necessary to prevent collisions with obstacles.

In general, the basic algorithm for define the nodal points of trajectory h(t) in the generalized coordinates space is very simple and does not require description. Note: the trajectory must be clarified (adjusted) at each nodal point of the motion.

When planning on a trajectory, the following constraints are applied [4]:

1. The cue points should be easily calculated in non-recurrence way.

2. The interim positions should be clearly defined.

3. It shall be ensured that the generalized coordinates and their two first derivatives are inseparable so that the planned trajectory in the generalized space is smooth (no excesses and leaps).

In general, the planning of a trajectory in Cartesian coordinates consists of the following sequential actions:

1. Formation the located along the planned motion trajectory the sequence of cue points in Cartesian space.

2. Select a class of functions that approximate the trajectory segments between the cue points according to some criterion (e.g. straight, arc of circle, parabola, etc.).

The first step gives you a high degree of accuracy along a given trajectory. However, it may require a large amount of computation that will slow down the manipulators movement due to recalculating the Cartesian coordinates into generalized coordinates if there are no sensors of the Cartesian coordinates of elements position. Therefore, the following action is to recalculate Cartesian coordinates of the cue points into the corresponding generalized coordinates, and interpolation by the polynomials of a low degree further.

IV. SIMULATING THE GEOMETRIC AND KINEMATIC PARAMETERS OF A MANIPULATOR

In order to program the manipulator motion in Autodesk Inventor, it is uses the auxiliary (transient) geometry of the Autodesk Inventor application programming interface (API). The auxiliary geometry is designed to solve by API a set of mathematical tasks, mostly geometric. Elements of auxiliary geometry are created using calls to the appropriate methods of the objects and are widely used in Autodesk Inventor API. Unlike most other geometric API objects, these objects do not have visual display tools. Most often, these are abstract mathematical proxy objects that are designed to manage the "real" geometry of Autodesk inventor.

To describe rotational and progressive connections between the neighboring links, Denavit and Hartenberg proposed a matrix method of sequentially building the coordinate systems associated with each link in the kinematic chain. The idea of the Denavit-Hartenberg presentation is to create a homogeneous transformation matrix that describes the position of each link's coordinate system relative to the previous link's coordinate system. It enables to consistently convert the coordinates of the manipulator elements from the last link coordinate system into the basic coordinate system, which is the inertial coordinate system for this dynamic system. Each coordinate system is based on the following three rules:

1. The z_{i-1} axis is directed along the i-th axis of the element.

2. Axis x_i is perpendicular to the z_{i-1} axis and directed from it.

3. The y_i axis supplements the x_i and z_i axes to the right Cartesian coordinate system.

The kinematic structure of the robot Fanuc LR Mate 200iC is presented in Figure 2.

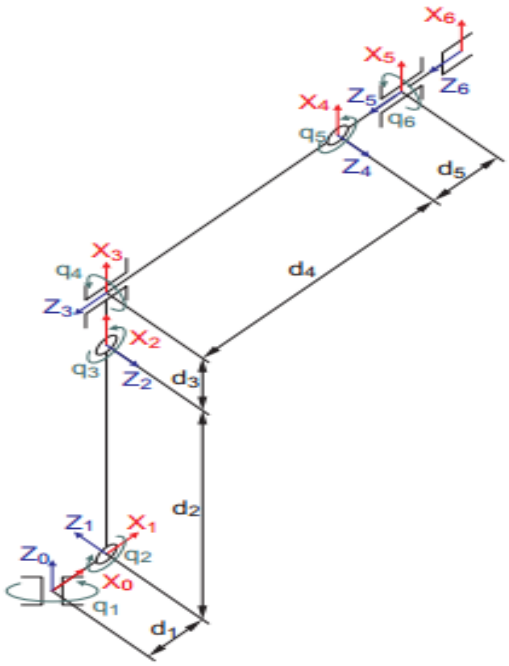


Figure 2. Kiinematic chain of robot-manipulator

The specified local coordinate system is associated with the i-th link. The configuration coordinates are specified through q_i. The Denavit-Hartenberg parameters are presented in Table 1.

The following geometric parameters of robot elements are used in Table 1:

d1 - The offset between the coordinates of x0, y0, z0 and x1, y1, z1 (mm).

d2 - Length of the first robot element (mm).

d3 - The offset between x2, y2, z2 and x3, y3, z3 (mm).

d4 - Length of the second robot Element (mm).

d5 - Length of the third robot element (mm).

TABLE I. DENAVIT-HARTENBERG PARAMETERS

№ link	a_i	α_i	d_i	θ_i
1	d_1	$-(\pi/2)$	0	q_1
2	d_2	π	0	$q_2-(\pi/2)$
3	d_3	$-(\pi/2)$	0	q_2+q_3
4	0	$\pi/2$	$-d_4$	q_4
5	0	$-(\pi/2)$	0	q_5
6	0	0	$-d_5$	q_6

V. MATHEMATICAL MODEL OF ROBOT-MANIPULATOR MOTION

To move a manipulator tool along a specified trajectory, a sequence of intermediate manipulator elements positions is specified. At the beginning, the coordinate arrays of these elements intermediate positions on the motion trajectory are defined. The next action is to define a set of rotation angles for all the manipulator elements for each coordinate on the trajectory. These coordinates must conform to the current position of the element on the trajectory (Fig. 3).

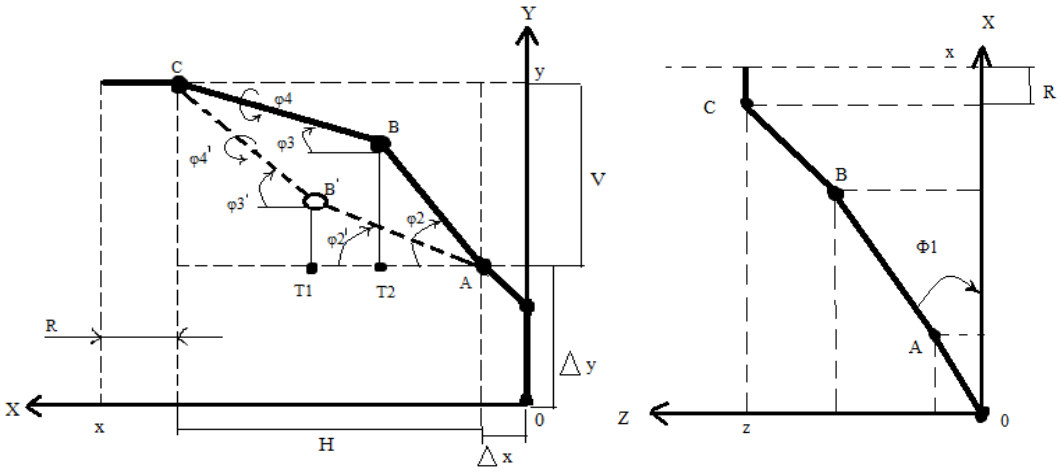


Figure 3. Rotation angle values in planes XY and XZ

For Example, to locate the point T, uses a classic task of crossing two circles solution:

$$T = K\ H \pm \sqrt{K^2 H - \left(V^2 + H^2\right)\left(K^2 - (AB)^2 V^2\right)} \Big/ \left(V^2 + H^2\right),$$

$$K = (AB)^2 - (BC)^2 + H^2 + B^2/2$$

There are two ways to implement model in Autodesk Inventor a controlled computer: to develop an application or to create a macro. It have chosen a method that consists of writing a macro by Visual Basic for application programming language, due to no additional software is required: Inventor has a built-in Visual Basic for Application language editor.

CONCLUSIONS

In the course of carrying out the work, an analysis of the existing information technologies for the control of robot-prototype computer models was carried out. The functioning controlled computer model of an industrial robot-manipulator and developed technology for its creation is a result. The application of the resulting technology in the education process enables to learn how to control robotic systems, to examine the functional requirements of the model development environment, and to reinforce the programming skills of the controller systems.

REFERENCES:

[1] I. Makarov, Y. Topcheev, *Robotics: history and perspective*. Moscow, Nauka, 2003. (In Russian)

[2] LR MATE 200IC INFORMATION [electronic source] URL: https://www.robots.com/fanuc/lr-mate-200iC

[3] V. Kontsevich, *Solid modeling in Autodesk Inventor*. Kyyv, DiaSoft, 2007. (In Russian)

[4] N. Svirnevsky, *Developing applications for Autodesk products*. Khmelnitsky, KNU, 2016. (In Russian)

SESSION 2
PROJECT MANAGEMENT

Cluster model of territorial community development program

Alona Starostina
O.M.Beketov National University of Urban Economy in Kharkiv, Ukraine, Starostina-2010@yandex.ua

Abstract. The mechanism of development program formation for territorial community based on the cluster approach has been offered. The mechanism is based on a conceptual cluster model of programs. This model includes four clusters: cluster of objects that form the capability of territorial community, cluster of objects that form "overcapability" of the community; cluster of objects that define the focus area of the territorial community; and a cluster of objects that do not meet the overall focus area of local community, but generate some profit. Based on different cluster layout in program, four versions of a mathematical model of the program have been developed.

Keywords: program, cluster, territorial community development, formation mechanism.

I. INTRODUCTION

For the moment, the program approach is increasingly used to provide clearly oriented development of territorial communities. This is evidenced by the implemented programs in the region. However, along with this, there is a certain "imbalance" of these programs' outcomes, both within territorial communities and within regions.

All this leads to a lack of so-called "capacity" of territorial communities, which is described in the Resolution of the Cabinet of Ministers of Ukraine No 214 of April 8, 2016 as the ability to "independently or through the appropriate local authorities to ensure the appropriate level of service, particularly in the field of education, culture, health care, social security, housing and utilities sector, given the human resources, financial support and infrastructure of an administrative-territorial unit" [1].

This makes the governing authorities of administrative units look for modern innovative approaches to managing the development of territorial communities that will ensure their "capacity".

After the conducted analysis, in terms of this issue, the author determined that to ensure sustainable development of territorial communities the most appropriate is to apply the cluster approach. And on this basis to improve program management methodology, particularly advisable is to develop a cluster, and on its basis, a mathematical program model of the territorial community.

II. MAIN BODY

In the general theory of cluster analysis, the term 'cluster' refers to a group of similar objects. Many scholars have paid the definition of 'cluster' considerable attention, including the following works of contemporaries [2 – 10]. Since, the management program of territorial community includes systemic and sustainable management of various projects and works, it is advisable to build a conceptual model of the program, which will reflect its structure and the contents. The conceptual model of territorial community development program, based on the cluster approach, is presented on Figure 1.

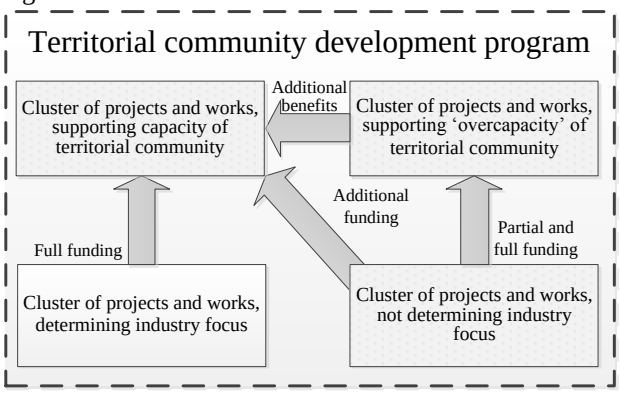


Figure 4. Conceptual cluster model of territorial community development program

According to the presented model of territorial community development program consists of four clusters, which consist of projects and works or so-called cluster objects. The first cluster includes objects that provide "capacity" of territorial community. This cluster includes projects implemented in enterprises, institutions and schools, designed to provide the appropriate level of quality of citizens' life. Moreover, quantitative characterization of the proper level should not exceed the limit set at the national level for similar performance indicators. First of all, public institutions, educational and medical institutions, and other entities that provide vitality of the region.Typically, projects of such a cluster are not profitable, meaning they need grants from other clusters of the territorial community. But along with this, the rejection of the projects of the cluster is not possible, as they are vital for the residents of the territorial community.

The second cluster includes projects that are designed to provide so-called "overcapacity" of territorial community. Meaning that projects and works, which create additional benefits for the residents of the territorial community, but is the sum of profit of territorial community from implementation of projects and works of third cluster, hrn. their absence does not have any critical impact on the life of the population. Such projects are also increasingly unprofitable and therefore can exist only in the case of sufficient funding. This cluster may be missing in the development program of the territorial community, the decision on its availability and feasibility should be made based on the presence of funding for implementation of these projects. By the projects implemented within this cluster include the so-called entertainment projects result in additional income for the local community, improving the indoor climate in the community. The lack of this kind of project is not critical to the functioning of local community as a whole.

The third cluster consists of projects that define the sectoral focus of the territorial community. That cluster now includes a group or group of objects that are most profitable for the territorial community. Besides this cluster projects are closely linked with the national strategic priority vectors, and should not contradict them, thus fulfilling the condition of the common development orientation of the territorial community and Ukraine. Profit from this cluster should provide the full funding of the requirements of the first cluster, providing thus independency of territorial community.

Also on investments of this cluster may fully or partly depend second cluster. The general position of the cluster is an important development program for the local community. as a matter of choice and Inclusion in the cluster is currently high. After all, it formed part of the program profitable.

Since, this cluster is the main sector in determining the direction of local community expedient to propose classification data on sectoral clusters feature. So the author proposed the following classification:

- agricultural cluster;
- heavy industry cluster;
- tourism and recreation cluster;
- building cluster;
- cluster of light industry;
- food industry cluster;
- innovation cluster;
- fuel cluster;
- energy cluster;
- machine-building cluster;
- wood-processing cluster.

Each of the proposed cluster has its own specific activities and needs of various resource supply, which should provide structure and programs of the local community.

The fourth cluster includes projects that do not meet the general branch orientation of territorial community, but bring a profit, which can be used to enhance the effectiveness of performance of the first cluster projects to a level higher than the one adopted at the state level, or for financing projects and works of the second cluster. Similarly to the second cluster, this cluster may be missing in the development program of the territorial community.

However, there are a development of the local community cluster indicates that the community has a number of potential areas of outsiders except those regulated cluster defining sectoral focus of the community. Therefore, the cluster management requires specific ratio, which is the continuous monitoring processes that occur in the community, making their analysis and if necessary corrective measures to local community development programs in general. Due to the fact that all programs of local communities are unique, ie different time implementation, scale, composition and structure, etc., the author believes is appropriate to highlight this issue in a separate scientific study.

A conceptual cluster model of development program of territorial community and the principles of its formation are the basis for building a mathematical model of the program. Given the fact that the program conceptual model shows clusters that can be present or absent in a particular program, it is appropriate to consider different options for application of mathematical models of territorial community development program, elaborated on the basis of cluster approach.

The first version of the mathematical model of territorial community development program is based on the assumption that in this program there are present only the first and third clusters. Therefore, the overall view of the program can be presented as the expression (1):

$$CK_1 \le PK_3, \tag{1}$$

where CK_1 is the sum of funding required for implementation of projects and works of the first cluster, hrn.;

PK_3 is the sum of profit of territorial community from implementation of projects and works of third cluster, hrn..

This model formalizes a situation in which profits from the objects of third cluster are sufficient for the fulfillment of capacity condition of the territorial community.

The second option of a development program mathematical model of territorial community is based on the assumption that there are first, third and fourth clusters. Its mathematical model has the form shown by (2):

$$CK_1 \le PK_3 + PK_4, \tag{2}$$

where PK_4 is the sum of profit of territorial community from implementation of projects and works of fourth cluster, hrn.

It is reasonable to note that this model formalizes a situation in which profits from the fourth cluster are directed to provide "overcapacity" of territorial community, and leads to improved quality of residents' life.

The third option of program mathematical model of territorial community is based on the assumption that there are first, second and third clusters, and at that the profitability of the last is at level sufficient to ensure the implementation of projects and works of the first and second clusters. A mathematical model has the form shown by (3):

$$CK_1 + CK_2 \le PK_3, \quad (3)$$

where CK_2 is the sum of funding required for implementation of projects and works of the second cluster, hrn.;

The last fourth option of program mathematical model is based on the assumption that there are all four clusters, and the model is expressed (4):

$$CK_1 + CK_2 \le PK_3 + PK_4. \quad (4)$$

The fourth model is usually present in large territorial communities, with a large population.

The choice of a mathematical model the most appropriate for each individual territorial community development programs depends on a combination of projects and works that are included. However, not always a simple division of projects and works into groups can result in an effective development program.

CONCLUSIONS

In the paper, there was developed a conceptual cluster model of territorial community development program, which includes four clusters: cluster of projects and activities that shape the capacity of the territorial community, a cluster of projects and works that form "overcapacity" of the community; cluster of projects and works that define industry focus of territorial community; and cluster of projects and works that do not meet the general branch orientation of territorial community, but generate some profit.Based on different cluster layout there are four versions of the program mathematical model developed. The classification of clusters that are outstanding for the application of the local community, on an industry basis.

Further research require: development of mathematical methods for assessing capacity and "overcapacity" of territorial community; development of profitability assessment of territorial community from projects and works of third and fourth clusters; development of mechanism for monitoring and control of dynamic processes in the local community environment that affects the development of the fourth cluster mechanism and introducing corrective measures to local community development program; the development of mechanisms for the most efficient distribution of funds between the components of clusters.

REFERENCES

[1] Postanova Kabinetu ministriv Ukrainy № 214 vid 8 kvitnia 2016 r. *Pro zatverdzhennia Metodyky formuvannia spromozhnykh terytorialnykh hromad* [On approval of methodologies capable of forming territorial communities]. Available at: http://zakon5.rada.gov.ua/laws/show/214-2015-%D0%BF. (accessed 11.04.2017)

[2] Kovalenko N.N. Theoretical approaches of foreign and ukrainian scientists to defining the concept "cluster", *Journal of Zaporizhzhya National University*, 2010, vol. 1(5), pp 11-19. (Теоретичні підходи зарубіжних та українських учених до визначення поняття «кластер»)

[3] Begun S. Using Cluster Analysis for the Research of the Demographic Situation in Regione, *Economic Journal of Lesia Ukrainka Eastern European National University*, 2016, vol. 2, pp.122-128. (Застосування кластерного аналізу для дослідження демографічної ситуації в регіоні)

[4] Lyashevska O. Cluster method are already methods of generating socio-economic development, *Clusters as a tool of regional development: scientific materials workshop (Feodosia, 16-20 July 2012)*, 2012, pp.37-39 (Кластерний метод в системі методів формування соціально-економічного розвитку регіону)

[5] Koroljuk J. G. Cluster analysis of regional systems in the context of state regulation of social and economic development. (Кластерний аналіз регіональних систем у контексті державного регулювання соціально-економічного розвитку) Available at: http://nbuv.gov.ua/portal/Soc_Gum/Dtr_du/2009_4/files/DU_04_09_Koroljuk_39.pdf.

[6] Ryadno O., Berkut O. A study of the structure and dynamics of differentiation of social and economic development of Ukraine based on a cluster analysis, *Economic herald of the Donbas*, 2016, vol. 1(43), pp.60-67 (Дослідження структури та динаміки диференціації соціально-економічного розвитку регіонів України на основі кластерного аналізу)

[7] Maliuga L., Zagorodniuk O., Zagorodniuk V. Cluster analysis of activity of the enterprises in the context of competitive environment, *Scientific Bulletin of Kherson State University*, 2016, vol. 17(1), pp.73-75 (Кластерний аналіз діяльності підприємств в конкурентному середовищі)

[8] Lyudmyla M. Ganushchak-Yefimenko. A conceptual approach to high-tech sectoralclusters development management within national economy, *Recent economic problems*, 2015, vol. 5 (167), pp. 112-116 (Концептуальний підхід до управління розвитком високотехнологічних галузевих кластерів в національному господарстві)

[9] Petruk G.V. Development of knowledge cluster definitions, *Vector Science TGU. Series: Economics and Management*, 2016, vol. 1(24), pp. 62-70 (Развитие понятийного аппарата знаниевого кластера)

[10] Pyrogov D. A. Conceptual principles of the regional socio-economic development based on economic clusters financial provision, *Electronic scientific edition "Efektyvna ekonomika"*, 2017, vol. 4. (Концептуальні засади фінансового забезпечення соціально-економічного розвитку регіонів на основі економічних кластерів).

SESSION 3
E-LEARNING TECHNOLOGIES

Remote laboratory: using Internet-of-Things (IoT) for E-learning

A.Kalashnikov, H.Zhang, J.Jennings, Misko (Myshko) Abramriuk
Sheffield Hallam University, UK, a.kalashnikov@shu.ac.uk, http://tinyurl.com/lokvnbt

***Abstract* – In order to equip engineering E-learners with realistic hands-on laboratory experience, remote access to an operating lab may be provided. Such an access can be implemented using live video streaming from a physical lab coupled with IoT telemetry and telecontrol.**

The progress of an ongoing remote laboratory project is presented. Three different video streaming solutions were compared side-by-side, the back end electronic architecture was fully developed and the most notable observations are now reported.

Keywords – remote laboratory, E-learning, Internet-of-Things, engineering education, low cost live video streaming.

I. INTRODUCTION

The relentless increase in human knowledge, together with the steady advances in technology, call for continuous improvements in the provision of higher education, especially in engineering, in order to make the best use of resources and keep the learners engaged with up-to-date educational activities. Modern online technologies provide, *i.a.*, rapid and inexpensive access to a range of useful learning resources such as vendor technical documentation and application notes, topical tutorials, forum discussions, educational videos and even complete massive open online courses (MOOC). E-learning aims to use the latest technological developments for reducing educational costs and widening participation. In the spirit of this goal, we have worked to combine the Internet-of-Things (IoT) and live video streaming technologies to create the 'presence effect' of an electronics engineering lab over the Internet.

It is important to differentiate between a virtual (i.e. simulated) and a remote (i.e. physical) laboratory environment. The former requires the development of one-off virtual reality or simulation software. Distribution and maintenance costs are negligible, but the development itself can be very costly if it is to feel realistic and engage the learners. By contrast, a remote lab gives online access to actual physical instruments and/or actuators and thus requires session scheduling and some maintenance. Some educational projects try to combine the two abovementioned approaches. One notable example is the India-wide Virtual Lab project [1], which, being mostly virtual, does include some provisions for 'remotely triggering an experiment in an actual lab'.

Wikipedia hosts a list of past and ongoing remote lab projects [2], although some of them no longer have an online presence. For example, the website for the Labshare project, which received AUD 3.8 million from the Australian government over 2009-2011 [3], is no longer being updated. We assume that the project could not continue because it was not possible to find long-term financing for its maintenance after the initial funding was used up.

Some of the present authors have been involved in the remote laboratory access project [4] that was accessed from national and international locations on numerous occasions and demonstrated at an IEEE profile conference [5]. It was dedicated to measuring the volt-ampere characteristics of a semiconductor diode. It was developed using an IP camera, which provided a live video stream, and an embedded web server to control the diode's voltage. Despite being a successful feasibility study, this project identified some serious usability shortcomings. These included:

- uneven natural lighting resulted in noticeable shadows and differences in image quality, depending on the time of the day [5, fig.4];
- it was impossible to host the setup at the university site because the corporate firewall for security reasons was configured to block all the external access attempts;
- the need to use two separate browsers (one for viewing the video stream and the other for control of the test rig) was inconvenient and resulted in noticeable lags between the diode's voltage changes and video response to these;
- national access was relatively smooth, but the international connections were smooth for twenty or so seconds, after which the learners' browsers needed to be restarted.

These issues were solvable in principle. For example, the last three could be addressed by employing a dedicated server with a video capture capability outside of the university's network. However, they could not be resolved within the remote laboratory's targeted budget.

This remote laboratory access project was restarted last year at Sheffield Hallam University, after the increased availability of single board computers (SBCs) such as Raspberry Pi and its competitors could provide decent computing, networking and video capture capabilities at low cost. Teaching enhancement funding was used to procure suitable hardware for the project. Undergraduate students of computer systems engineering were tasked to set up and test several alternatives for live video streaming as part of their second year project module. At this stage we aimed to develop the back end IoT electronic architecture and compare different ways of

streaming live video. This paper presents a snapshot of our development results and findings to date.

II. WHY THE VOLT-AMPERE CHARACTERISTIC OF A SEMICONDUCTOR DIODE?

The laboratory experiment, accessed remotely, must satisfy various criteria in order to be genuinely useful to the learners. First, it should fit well into the universally-accepted curriculum of a relevant course. Second, it should use common industry-standard equipment to achieve the presence effect. Third, it should have a limited number of sensors and actuators that can be easily controlled and/or read by the back end electronics. This list can be expanded by adding various cost (development, maintenance, running), pedagogical, safety and other requirements. We selected measurement of the volt-ampere characteristic of a semiconductor diode for the following educational reasons:

- this is likely the first semiconductor device that electrical and electronics students encounter in their studies;
- measurements require two digital multimeters, which only a few learners have access to for conducting such an experiment on their own;
- there are many types of semiconductor diodes which have significant differences in their characteristics (Si and Ge based power and small signal standard diodes; Schottky, Zener, tunnel and Gunn diodes to name a few, some of which are difficult to source at low cost).

There are also some economic reasons:

- the setup requires only two standard multimeters and a regulated variable voltage power supply;
- automation enables keeping the diode's current within safe limits, eliminating the possibility of blowing it up during experimentation or by connecting it incorrectly, thus lowering maintenance costs and keeping the learner safe.

III. WHAT ARE THE OBJECTIVES OF THE PRESENT DEVELOPOMENT?

We plan to develop and deploy a remote access solution for an existing undergraduate lab experiment: measurement of a diode's volt-ampere characteristic. The students will be encouraged to use the online version or asked to use both and describe their user experiences. The learners' feedback will be collected and analysed in order to enhance the hardware/software/firmware provisions and lab sheets. Most of the developments are to be conducted as part of the academic assignments for various project modules.

IV. LOW COST OPTIONS FOR STREAMING LIVE VIDEO AND OUTCOMES OF THEIR TRIALS

We originally anticipated that using a Raspberry Pi-like SBC, which is built around a complete system-on-chip integrated circuit to reduce costs, would be the only feasible option for the project. Raspberry Pi seems to be better supported than its alternatives, and the latest Raspberry Pi 3 is the most capable SBC of the RPi family at present. However, a recently-launched Intel Compute Stick (a complete small factor PC with an HDMI output) [6] made an unexpected contender to most ARM-based SBCs, and its inexpensive alternative MeeGoPad T07 [7] was bought as an alternative to RPi. It was used on Windows 10 to explore non-Linux options for the project. Affordable video capture options at present include:

- IP cameras (complete video streaming solution which only requires a network connection to operate; built-in protocols allow bypassing simple Internet firewalls, which is a security concern but could simplify access for this project; highest cost; was used in the previous development);
- web cameras (connected via USB to a host computer which should feature a streaming video web server; moderate cost; could be used for both computing platforms);
- cameras which connect directly to the dedicated SBC pins (the lowest cost if available; only the RPi has this option available).

Out of the five usable SBC/video capture combinations, the following three were explored by separate groups of students (fig. 1):

- Raspberry Pi + IP camera;
- Raspberry Pi + RPi camera;
- MeeGoPad T07 + web camera.

The students' trials resulted in the following findings. All the groups managed to enable video streaming over the same LAN. A wired LAN connection was found to be much better than the WiFi one, even for the MeeGoPad that was connected to a wired LAN via an additional USB-to-Ethernet bridge. (It should be noted that the same connection is internally used in RPi as well.) The MeeGoPad demonstrated the best performance in terms of image quality and responsiveness. Access to the firewalled streaming video server via a virtual private network was found to be cumbersome, and too many access rights had to be given to an occasional user. When NGINX and Apache web servers were compared against the RPi with the RPi camera, the former demonstrated much better performance. Custom web pages, developed to display the video stream with some user controls (Fig.1), required combined use of HTML, CSS and JavaScript.

V. THE BACK END ELECTRONIC ARCHITECTURE

The architecture was devised to address the following principal requirements:

- parts to be used must be mass-produced, inexpensive, easily obtainable and swappable;
- they must conform to relevant safety standards, operate unattended and have either low power

consumption or be capable of operation in low power mode to save on running costs;

- the number of custom devices should be kept to a minimum.

The setup is to be powered from a standard laptop-like power supply. These are inexpensive, can safely be plugged into the mains for a long time, and waste little power in stand-by mode. Laptop power supplies output a fixed-value, fixed-polarity voltage of around 20V. To extend the applicable voltage range and enable its smooth control from negative to positive values without noticeable cross-over distortion, a boost converter with fixed output voltage will power two high-voltage operating amplifiers (opamps). Such converters are available from many manufacturers, provide reasonable conversion efficiencies and keep their set output voltages reasonably stable across a range of temperatures.

The output voltage of the first opamp, used as the voltage follower, is fixed at the mid-point of its supply voltage by connecting its non-inverting input to a resistive potential divider, which halves the power supply voltage to create the virtual ground. The second opamp, used as an inverting amplifier with a suitable gain, has its non-inverting input connected to the output of the first opamp, and its non-inverting input connected to the variable voltage source, which is controlled over the Internet. The diode in question is to be connected between the opamps' outputs (bridge configuration), and the voltage across it will always be equal to the remotely set voltage times the second opamp gain, irrespective of any supply voltage fluctuations. Dual channel high-voltage opamp KA334 with built-in thermal shutdown was selected for added assurance of safe unattended operation. The complete custom electronic circuit requires fewer than 10 electronic components in total.

The previous development used a single hardwired diode, which limited application of the setup to a single experiment per learner. For this project, an additional board will be used to select one diode out of the available set either at random (assessment mode) or by user selection (exploration mode). A standard multi-channel electromechanical relay board is to be used for this purpose. The diodes will be connected to the relay board using a dummy board with screw terminals to enable quick replacement of the tested parts, which may include other non-linear two-terminal components such as resettable fuses or voltage surge protectors. If an RPi with its exposed pins is used as the web server, a standard Pi HAT input/output board can control the relay board, and a standard Pi HAT DAC board can control the voltage applied to the diode. If a MeeGoPad is used as the server, there will be a need for an additional microcontroller board connected to it via USB. An Arduino-compatible board with an additional DAC shield seems adequate for this task.

VI. Further project goals

This summer the back end architecture will be implemented and tested. At the start of the 2017/18 academic year, new project students will be tasked to integrate both the video feed and controls into one web page. If possible, we will find solutions for connecting to the remote lab from outside the corporate firewall, and road test the performance of the remote lab with the view to introducing the lab to the curriculum in the academic year of 2018/19.

VII. Conclusions

Although the project is still far from being completed, we consider the following results most useful for further development:

- the performance of the most recent Raspberry Pi 3 for live video streaming was inferior to that of the MeeGoPad with a web camera;
- live video streaming using wired LAN connections significantly outperformed those using built-in WiFi adapters;
- design of a modern responsive front end for the remote lab will necessarily require some expertise in a combined use of HTML, CSS and JavaScript;
- electronic back end architecture was fully designed and will be implemented over the summer to be used along with next year's student project activities;
- a solution for hosting the remote lab behind a corporate firewall is yet to be found.

Acknowledgments

The authors gratefully acknowledge support received from Nottingham (Lord Dearing fellowship, 2010) and Sheffield Hallam (grant from the teaching enhancement fund, 2015) universities. The following students, listed alphabetically, contributed to the project: B.Al Arayedh, S.Carswell, A.Gray, A.Hussain, A.Kahin, A.Maggs, J.Mercer, D.Parkin.

References

[1] Virtual Labs (India), abvailable online on tinyurl.com/kfg82v3, accessed April 2017.
[2] Remote laboratory, available online on tinyurl.com/mowejgw, accessed April 2017.
[3] Labshare, available online on tinyurl.com/l2jra5b, accessed April 2017.
[4] Remote laboratory acess project, available online on ak2015.uk/b/, acessed April 2017.
[5] E. Otoakhia *et al*., "Embedded web server for remote laboratory access for undergraduate students studying electronic engineering," *2011 IEEE International Symposium of Circuits and Systems (ISCAS)*, Rio de Janeiro, 2011, pp. 337-340. doi: 10.1109/ISCAS.2011.5937571
[6] Intel Compute Stick, available online on tinyurl.com/nrmx5n2, accessed April 2017.
[7] MeeGoPad T07, available online on tinyurl.com/llg499f, acessed April 2017.

Fig.1. Hardware used by different student groups (top row) and corresponding screenshots (bottom row)

Problems of distance learning systems monitoring and evaluation of their efficiency

Oleksandr Viunenko [1], Andrii Tolbatov [1], Volodymyr Tolbatov [2]
[1]Sumy National Agrarian University, [2]Sumy state university, tolbatov@ukr.net

Abstract – The article is devoted to problems of high tech technologies implementation in the information-intensive educational process in the university.

***Keywords* – distance learning, distance learning management systems, information and communication technologies.**

I. INTRODUCTION

Practical implementation of monitoring models of innovations′ efficiency points at a number of conflicts between current "stereotype" descriptive approach to efficiency evaluation of innovations and a need to discover criteria and factors, which allow to correct evaluation of the innovation implementation efficiency, as well as between traditional educational institutions with regular educative practices and a need from the state and society to use adequate systems to evaluate the implemented innovations. These conflicts determined the problem of the investigation. The problem involves a need to establish the content of monitoring models of innovations′ efficiency in the distance learning systems (further - DLS), peculiarities of their elements, criteria and factors [1 – 12].

II. II. RESULTS

New legal background in Ukraine activated the development of distance postgraduate education. But, from the organization and economic point of view, more important is the professional development without discontinuing work. In addition, one more advantage of distance learning is an effective use of intellectual and creative resources of academic and teaching staff, possibility to create modern academic programs, which are constantly updated.

There are new opportunities, based on crowdsourcing and noosoursing, popular in the professional sphere, and priorities, concerning creation of the sphere, where constant knowledge generation becomes a vital need of professionals, are determined.

The implementation of innovative educational technologies is one of the few ways to update content and form of educational process organization, to provide equal access to modern high-quality education, and, in general, to actualize national doctrine of education development.

The factors, which affect negatively the organization of innovation activity in higher educational establishments, and shortcomings in activity of educational authorities, should be also noted:

1) absence of special-purpose investments in the innovation processes;

2) an insignificant level of theoretical and scientifically methodical qualification of the staff;

3) an insignificant level of activity of regional psychological services for psychological follow-up of innovation processes;

4) an insignificant level of educational-material base of universities;

5) periodic behavior of monitoring studies of innovation activity of educational institutions;

6) an insignificant level of public awareness campaign among applicants and specialists of modern educational innovations;

7) lack of cooperation between the universities while testing and implementing educational innovations;

8) problems of academic support of innovative educational activities from academic organizations and field scientific organizations;

9) no regulatory and legal framework, which can confirm the status of experimental educational institutions;

10) insufficient introduction of the administrative educational innovations in the activity of education authorities and heads of educational institutions.

The development of the structure of the DLS monitoring systems should include the following stages: 1) definition of the monitoring goals and objectives; 2) definition of the object and subject of the DLS monitoring; 3) definition of the monitoring sources; 4) development of the information-gathering methods; 5) definition of monitoring materials′ structure and content; 6) definition of the monitoring procedure.

The object of the DLS monitoring is an innovative reformation, while the subject is efficiency evaluation of innovations implementation.

The monitoring sources are:

a) documents, confirming using of innovations;

b) survey results, questionnaire surveys, interviewing etc.;

c) focus-groups material;

d) analysis of publications in the fields of innovation development.

Therefore innovations monitoring system will allow to:

1) perform quantitative and qualitative analysis of the development and implementation processes of innovations;

2) analyze efficiency of innovative changes;

3) discover factors, preventing the complete implementation of innovations, "tension" elements in the system;

4) determine the dynamics of quantitative and qualitative indicators of the process efficiency of innovative reformation.

The most important stage for building the monitoring information system is to determine principles of system organization and evaluation of its efficiency. The existent monitoring analytical systems are mostly focused on the staff cost minimization, planning and technologic forecasting of scientific research and development, analysis of the activity results. To create effective monitoring information analysis system, it is necessary to review the main principles of building the basic model [3,4,5]:

1) to create individual database register of employees, accounting methods and analysis processes;

2) to develop the algorithm of efficiency and potential evaluation of the scientific activity;

3) to create possibility for monitoring on a regular basis.

The system information base should also collect the data in the following areas:

1) the publication′s activity and citation rating of scientific works, their citation index;

2) state training system for academic and teaching staff;

3) international activity, foreign grants;

4) participation of academic and teaching staff in economic contracts, grants and target programs.

Change of any of these elements can change other dependent factors. That is why, while building the system, it is necessary to take into account the combined effect of external and internal factors. Consequently analytical system of DLS monitoring should contain informative and control components, use the organization potential effectively, and storage, data-processing and control functions for rational decision-making [6].

It should be also considered that modern technologies in the information society always promote the virtual environment for the individual, who lives in the real society. A modern person manages not material objects, but symbols, ideas, images, intelligence, knowledge. It means that there is a need to solve informational and ecologic problems, problems, concerning information security of an individual and society, information wars, collective consciousness manipulation, computer-related crimes, virtualization of interpersonal communication, protection of the human nature in a context of creating artificial intelligence etc [7].

Distance learning system as a social organization and technology is based on the practical use of knowledge about interaction peculiarities between human and info-communication technologies for continuous education and self education, use of interactive technologies for organization of virtual educational process. Info-communication technology is an essential characteristic of modern educational process, while high educational technology is an inherent characteristic of an effective distance learning.

Principles of integrity, structure and additionality as methodological principles of educational process development also allow to solve the conflict, which arises because of incompatibility between traditional approaches and new educational technologies; to determine the development paths of the educational process structure; to combine traditional and innovative components of educational process into one integral distance learning system; to unite instuctivism and constructivism strategies in education; to overcome representation conflicts of members of innovative educational activities; to generalize the existent types of communicative competence; to determine perspective of cognitive learning.

One of main factors of distance learning development are the following: a need to train professionals and intellectuals, creative specialists, who are able to coordinate, accumulate and convey knowledge; an active search of the improvement mechanism for institutional and management structure; positive consequences of globalization, internationalization and informational support of education that allow to combine technocratic trends, beliefs and academic culture.

Traditional form of learning process organization, which involves live interactive communication between the lecturer and the student, dominates for educational needs of students. But the complete model of educational system is still in demand. It depends on the fact that the requirement level for information and communication technologies usage in educational process is higher than the real student knowledge and skills, and because distance learning is a new mode of study for higher education without wide practical use.

It should be also considered that, except a huge number of software programs represented as super-tutors, prof-tutors, complays, R-tutors etc., a complete educational model suggests a continuous use of such types of telecommunications as teleconferences, tele-tutorings, various imprinting video films and slide lections, which require heavy financial expenses.

Distance learning system actualizes characteristics of virtual reality. However, distance learning appeared as a new form of extramural studies. That is why it is closely connected with intra-extramural and extramural form of study.

At the same time distance learning is an alternative to traditional extramural learning, however, distance

learning technologies are actively used in traditional intramural studies. They provide more effective introduction of information and communication technologies. It means that in distance learning the student is the active subject, and the percent of creative independent work is constantly increasing.

Lately the role of higher education institutions in building regional information systems and advisory services has increased. It is connected with the development of the following directions [8]:

- - reproduction of intellectual potential, necessary for knowledge development and commercialization;
- - production of innovative products and services by own efforts;
- - incubation and generation of research-intensive small enterprises;
- - forming of the support infrastructure of scientific researches, which serve with the needs of regional innovation systems;
- - development of innovative culture in business environment.

The proactive attitude of higher educational institutions in the development of the above mentioned kinds of activities will allow forming innovation activity centers for regions and other individual fields. However, it is impossible to prioritize the scientific and innovative activity of the higher educational institution, as a part of social and economic environment, without solving a pair of high-priority management tasks:

1) conclusive and complete identification of the object of management;

2) creation of an indicator system, evaluation factors and ways of their determination as a base to receive the information to make well-founded decisions concerning the innovative strategies development of higher educational institutions.

By now these tasks are not solved, it complicates the development of scientific suggestions and decisions, concerning the activation of innovation activity of higher education. The innovative potential, accumulated by universities, needs to be structured, structurally analyzed and monitored.

While solving the tasks, concerning the development of in-house infrastructure, the following problems of monitoring and evaluation of universities efficiency have the top priority:

- to bring the innovation activity in the line with scientific and educational activities by creating innovation system of higher institutions with the main institutional elements, represented by innovative and active units - academic departments, centers, institutes, laboratories etc.
- to form the ongoing institutions, necessary to market the science-intensive products (exhibitions, contests and other interactive technologies), and associative communities to protect interests of innovative business, to develop material and technical base of universities;
- to create the network of small science-intensive enterprises on the basis of, or assisted by, higher educational institutions, aimed at commercialization and use of scientific potential of higher education;
- to create an infrastructure of university innovative management, basing on specialized subdivisions of transferring technology and intellectual property, created on their base, together with businesses and intellectual recourses;
- to create industrial research laboratories and research centers, together with large regional and national corporations, that will be included in the innovation processes of technological clusters and enterprises of real sector of economy;
- to deliver the innovation activity results to academic and scientific processes, aimed at extended reproduction.

Modern monitoring systems are created in the educational institution for targeted investigation, continuous monitoring of academic process and diagnostics of its quality by means of information analysis and special researches and measurements.

The monitoring result is events planning, concerning the improvement of educational quality and efficiency by delivering true evaluation of the academic process organization in the distance learning system.

The monitoring system itself is a mean to manage expenses by setting them into one-to-one correspondence with the received educational result.

Now the progressive are social monitoring methods and investigation of continuous educational systems with the use of distance technology of interconnection with consumer needs and regional social and economic environment.

They allow to predict possible results, to implement corrective measures in the individual student′s training, to guarantee a necessary quantity and efficiency of education by means of improving management and using innovative technologies.

Therefore, the building of effective information model of management becomes actual. It will allow to plan and organize the academic process including distance learning both on operative, tactical, and strategic levels of the university development. Only single information model of universities system will allow solving the following tasks:

- to create a structure for planning and monitoring the implementation of education program;
- to plan implementation of education programs on faculties and institutions;

- to organize monitoring and business-analysis of the implementation of education program on the basis of the credit-module education system;
- to optimize the implementation of education programs basing on requirements of business-processes;
- to plan and implement information model to organize various types of education for various educational levels, including distance learning
- to develop common functional requirements to develop and improve the information system of university.

Eventually, the most relevant task of managing scientific and innovation potential of universities is the development of formal evaluation procedures and development of information base with attached for rational decision making functions of storing, accumulating and processing data and management.

The main direction to solve the tasks of distance learning system development is improving of the employees′ competence, research skills, studying new information technologies, designing and analyzing business-processes, aimed at the higher result of the activity of the members.

The important role in this process is scientific justification of practical recommendations, concerning the information support of monitoring system, and reliable operative scientific and information support of management processes of the university scientific potential.

III. CONCLUSIONS

In the nearest future the creation of well-equipped institutions of distance professional development will become relevant and economically justified. They will give an opportunity to use intellectual and creative potential effectively by decreasing expenses for technical equipment and computer software, but to greater extent, expenses of state and private establishments for professional development. Refocusing on the working people, who have a certain motivation and long-term project of life, who are able to build their own development path, and who understand, that their purposes need certain resources. It can activate the development of the distance learning in the industrial and regional aspects.

REFERENCES:

[1] Viunenko O.B. Information technologies in the educational process as the basis of modern distance learning / Oleksandr Viunenko, Andrii Tolbatov, Svitlana Vyganyaylo, Volodymyr Tolbatov, Svitlana Agadzhanova, Sergii Tolbatov // TCSET 2016 – Lviv-Slavske, 2016. – P. 831–833.

[2] Tolbatov A. V. Rozrobka ta pidtrimka intelektualnoyi sistemi distantsiynogo navchannya u VNZ / A. V. Tolbatov, V. A. Tolbatov, S. V. Tolbatov, D. I. Chechetov // Perspektivnye innovatsii v nauke, obrazovanii, proizvodstve i transporte ‘2013: sb. nauch. Tr. Sworld. – Ivanovo, 2013. – Vyp. 4 (13). – S. 18–22.

[3] Brooking A. Intellectual Capital: Core asset for the third millennium Paperback – July 30, 1996.

[4] Makarkin N.P., Cherkasov V.D., Fadeeva I.M., Makolov V.I. Razvitie issledovatelskogo potentsiala v sisteme strategicheskih prioritetov universiteta. Universitetskoe upravlenie 2008;(6):46-51.

[5] V‘yunenko O.B. Pobudova sy`stem monitory`ngu, analizu ta ocinky` pry`jnyattya rishen` regional`nogo rivnya dlya sy`tuacijny`x centriv APK / O.B. V‘yunenko, A.V. Tolbatov, S.V. Agadzhanova, V.A. Tolbatov, O.B. Shandy`ba, S.V. Tolbatov // Vy`miryuval`na ta obchy`slyuval`na texnika v texnologichny`x procesax. – Xmel`ny`cz`ky`j, 2015. – №4. – S.194-201.

[6] Tolbatov A.V. Modern technologies of distance learning in agrarian higher school / S.V. Ahadzhanova, K.H. Ahadzhanov-Gonsales, A.V. Tolbatov, O.I. Zorenko, V.H. Lohvinenko, N.L. Barchenko, V.A. Tolbatov, S.V. Tolbatov // SW Journal Pedagogy, Psychology and Sociology. Scientific world, Ivanovo, 2015. – Volume J21508 (9). [November 2015]. – P. 109–114. – URL: http://www.sworld.com.ua/e-journal/j21508.pdf

[7] Tolbatov A.V. Dosvid vprovadzhennya texnologij dy`stancijnogo navchannya u vy`shhy`j agrarnij shkoli / A.V. Tolbatov, V.A.Tolbatov, S.V. Agadzhanova, K.X. Agadzhanov-Gonsales, N.L. Barchenko, O.Y`. Zorenko, V.G. Logvinenko, S.V. Tolbatov / Innovatsionnyie podhodyi k razvitiyu obrazovaniya i vospitaniya. – Odessa: KUPRIENKO SV, 2015. Glava 3. – S. 45–59.

[8] Tolbatov A.V. Funkciyi, osnovni skladovi ta osobly`vosti monitory`ngu dy`stancijnoyi osvity` v VNZ / S. V. Agadzhanova, O. B. V'yunenko, A. V. Tolbatov, K. X. Agadzhanov-Gonsales, V. A. Tolbatov // Novitni komp'yuterni texnologiyi. – Kry`vy`j Rig : Vy`davny`chy`j centr DVNZ «Kry`voriz`ky`j nacional`ny`j universy`tet», 2017. – Tom XV – S. 131–139.

[9] Tolbatov A.V. Modelyuvannya v osviti: Stan. Problemy`. Perspekty`vy` : Monografiya. Aktual`ni py`tannya pobudovy` sy`stemy` monitory`ngu dy`stancijnoyi osvity` agrarny`x VNZ / Tolbatov A.V., V'yunenko O.B., Agadzhanova S.V., Tolbatov V.A., Agadzhanov-Gonsales K. X. (Rozdil 2.6) Za zag.red. Solovjova V.M. - Cherkasy`: Brama, vy`davecz` Vovchok O.Yu., 2017. - 266 s.

[10] Tolbatov A. Data Representing and Processing in Expert Information System of Professional Activity Analysis [Text] / Oleh Zaritskiy, Petro Pavlenko, Andrii Tolbatov // TCSET 2016 – Lviv-Slavske, 2016. – P. 718–720.

[11] Tolbatov A. Mathematical models for the distribution of functions between the operators of the computer-integrated flexible manufacturing systems [Text] / Evgeniy Lavrov, Nadiia Pasko, Anna Krivodub, Andrii Tolbatov // TCSET 2016 – Lviv-Slavske, 2016. – P. 72–75.

[12] Tolbatov A.V. Information technology for data exchange between production purpose integrated automated systems [Текст] / P.M. Pavlenko, A.V. Tolbatov, V.V. Tretiak, S.V. Tolbatov, V.A. Tolbatov, H.A. Smolyarov, O.B. Viunetko // Vy`miryuval`na ta obchy`slyuval`na texnika v texnologichny`x procesax. – Xmel`ny`cz`ky`j, 2016. – №1 –S.86–89.

The comprehensive research of students' preferences in the electronic educational environment of the high school

N. Barchenko, senior lecturer [1]; E. Kaba, student[2]; E. Lavrov, Ph.D. professor[2];
[1]Sumy National Agrarian University
[2]Sumy State University,
prof_lavrov@mail.ru

The problems of user modelling are described. The system for evaluation user's preferences is developed.

Keywords. Ergonomics, Human-Computer Interaction, user.

I. INTRODUCTION

Ergonomic support of e-learning allows to improve the efficiency of the dialog interaction in the "student-to-computer" system.

The account of user`s preferences in developing interface and dialog procedures is significantly increase the value of user`s comfort and reduces the number of failures from activities when interacting with the system. [1-3].

In [2-3] has been developed the concept of an agent-manager of ergonomic support of e-learning.

The topical task is determination of user's preference and individual psychophysiological characteristics.

II. BASIC RESEARCH

2.1. Basic statements

To solve the problem of determining user preferences and psychophysiological characteristics, a subsystem was developed with the next main functions:

- Testing the student's psychophysiological characteristics;
- Determination of the student's preferences through the presentation of fragments of educational material of different modalities;
- Maintaining a database of test results;
- -Preparing reports based on the results of testing.

2.2. The forming of psychophysiological component (PPC) of user`s model

In [4, 5] are determined the psychophysiological characteristics of the user, which have a significant effect on the effectiveness of human-machine interaction during the learning process in the student-computer system.

They include:

- modality;
- temperament;
- level of endurance of the nervous system;
- type of nervous system;
- neuroticism;
- fatigue, etc.

Test methods have been defined that allow quickly obtain the values of the above described parameters of the PPC user`s model.

The main requirements for test methods are time and ease of execution.

2.3. The subsystem of forming PPC user`s model.

At present in subsystem were created the next test methodology:

- VAK-questionnaire – the determination of the preferred modality;
- The definition of preferences by viewing samples of educational material of different modalities;
- A Raven Matrices Test - determination of the level of intelligence;
- Eysenck Personality Questionnaire – the definition of temperament.

2.4. The use of testing subsystem

Knowing the dominant channel of information perception training can be made more effective [4].

15 students of the Sumy State University took part in the experiment. Results of answer interpretation for preference defining are presented on Figure 1.

V- Visual preferences.

A- Audio preferences.

R – Reading preferences.

K – Kinesthetic preferences.

Processing of VARK-questionnaire results showed that prevail students with mixed modality. The vast majority of students have prevailing audiological modality.

Results of answer interpretation are presented on Figure 2.

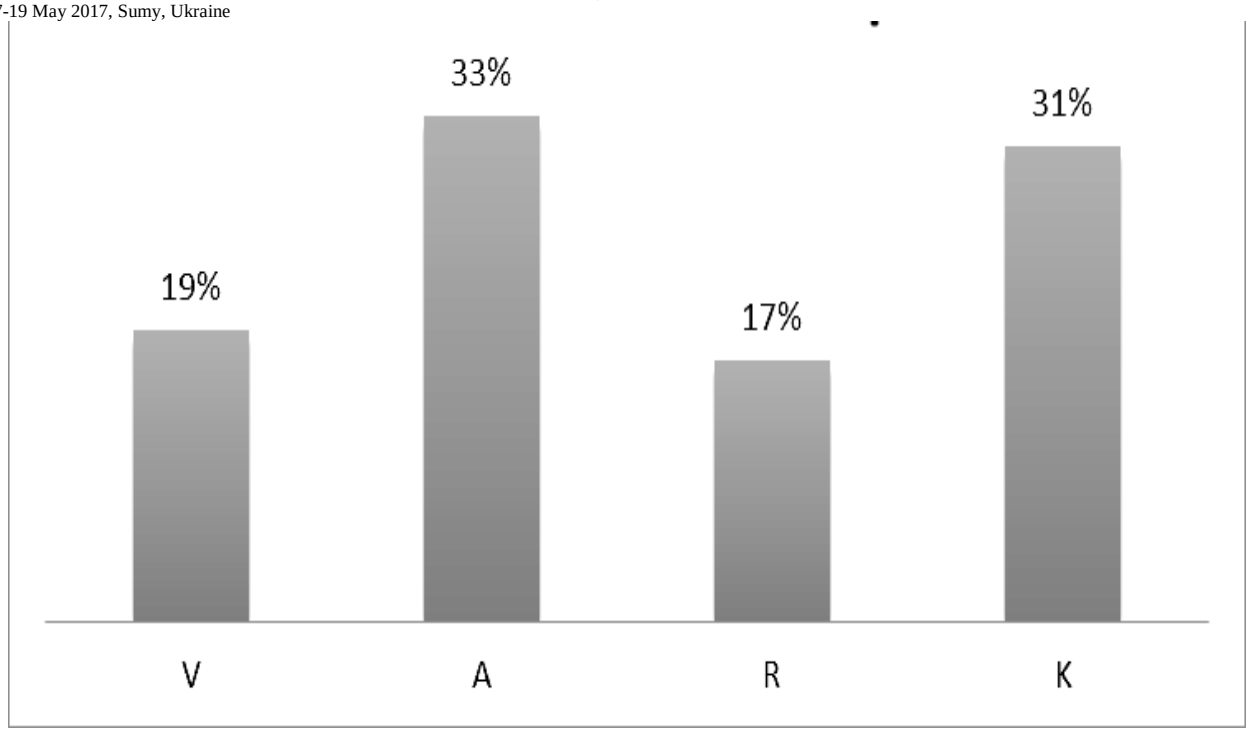


Fig.1 Video chart of the results of testing user's personal preferences

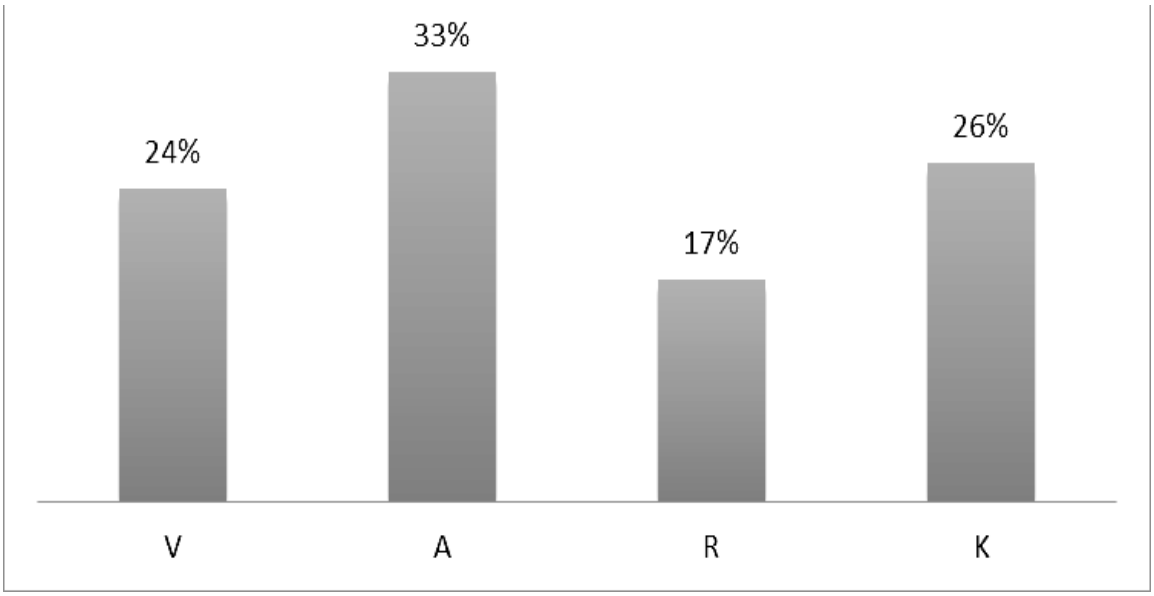


Fig.2 Video chart of the results of VARK testing

The results of the answers interpretation of the Eysenck Personality Questionnaire for determining the temperament are shown in Figure 3. There are four types of temperament: choleric, sanguine, phlegmatic, and melancholic. As a result of testing among students the phlegmatic type of temperament prevails.

Fig.3 Video chart of the results of Eysenck Personality Questionnaire

Among the 5 degrees of intelligence, the vast majority among the students tested has a third degree. The results of the interpretation of the responses to Raven Matrices Test to determine the level of Intelligence are shown in Figure 4.

Interpretation of degrees is as follows:

1 degree - very above average (subject has highly developed intellect);
2 degree - above average;
3 degree - the average intellect;
4 degree - the intellect is below average;
5 degree - defective intellectual ability.

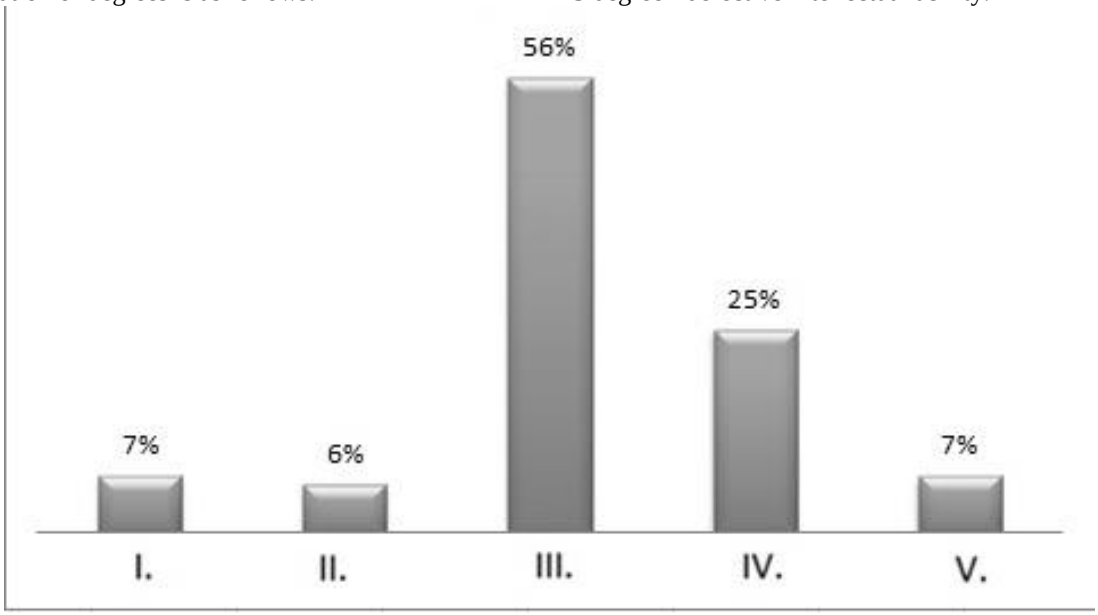


Fig.4 Video chart of the results of a Raven Matrices Test

III. THE DIRECTIONS OF FURTHER STUDY

To form a user model, it is advisable to use a test system that includes: modality, a Raven Matrices Test, Eysenck Personality Questionnaire.

The next stage of study is an integration into the real learning systems.

REFERENCES

[1] Lavrov E., Kupenko O., Lavryk T., Barchenko N. Organizational Approach to the Ergonomic Examination of E-Learning Modules //Informatics in education , 2013, Vol. 12, No. 1 - 105-123.

[2] Lavrov, E.A. Podhod k sozdaniju intellektual'nogo agenta dlja sistemy jergonomicheskogo obespechenija jelektronnogo obuchenija [Tekst] / E.A. Lavrov, N.L. Barchenko // Suchasni informacijni sistemi i tehnologiї : materiali Drugoї mizhnarodnoї naukovo-praktichnoї konferenciї— Sumi : SumDU, 2013. — S. 93-94.

[3] Lavrov, E.A. Agent-menedzher v sisteme jergonomicheskogo obespechenija jelektronnogo obuchenija [Tekst]/ E.A. Lavrov, N.L. //Bionika intellekta. – 2013. – №2 (81). – S. 115–120

[4] Fleming, N. D., & Mills, C. Not another inventory, rather a catalyst for reflection. 1992

The methods and processes of creating virtual simulator

N. Fedotova, M. Krishtop, T. Jabarin, M. Shelikhov,
Sumy State University, Computer Science department, Sumy, Ukraine, fna_2000@ukr.net,

Abstract – Actuality development of multimedia models of arms and military equipment due to both scientific purposes theoretical concepts and practical purposes a more effective educational systems (systems) and their implementation should ensure the necessary quality of military education, in particular accordance with the documents governing the content of education.

***Keywords – information technology; conceptual mode; CMSA; virtual simulator*.**

I. RESEARCH METHODS

Identification of the idea of the project is to develop a multimedia complex computerized training of military specialists.

The problem is the lack of educational materials, increased danger at training on a real technique, great price for the purchase and maintenance of educational layouts a real technology.

Purpose is to improve the quality of learning.

The aim is to develop a computerized method of learning based on the three-dimensional graphics.

Process of creating a virtual simulator is used methods of simulation and performed several working stages:

1. Learning of physics processes input and installation of measurable parameters. This stage of work is determined, of which the main elements will be built imitation of the shutter mechanism. Knowing the specific input parameters, decide which method will be implemented controls virtual model - "devices" regulation.
2. Create geometric models shutter mechanisms. At this level, the virtual model graphics solution - advanced virtual simulators performed in three-dimensional graphics with a maximum imitation of materials and lighting that significantly improves the quality of work.

The main objective here is to approximate the model to the real object, through compliance with correct proportions, sizes, colors and lighting [1].

3. Development of interactive modules, combining geometric patterns and physical dependence. Writing code virtual model is the most time-consuming part of the job. The task is to develop an algorithm that adequately describes the physics of the interaction mechanisms of the shutter.

The program binds together the graphics, sound and text support, interactive component, and according to precise mathematical relationships, simulates the dynamics of the process or phenomenon [2].

4. Implementation of guidance and background information. When the virtual model is generated, it is necessary to provide accompanying methodological information or informative, allowing the user to more fully explore the nature of the study and master the management of a virtual simulator.

At this stage, the major challenge is structuring all educational material to make affordable, easy to handle "workplace" that teaches the user. One effective way to implement the system of methodological support virtual simulator is to develop a program environment that allows the user to get acquainted with the structure of the training course, to make direct access to sections of the course, run interactive modules labs, store and read statistics passage rate [4].

After following steps to create a virtual simulator developed information technology scheme, which is shown at the Figure 1.

Creation of an information system anyway faces with a problem of information exchange. Information exchange – a process which consists of the sender and recipient who are connected by a reliable channel of communication that is the message from the sender to the recipient using the communication channel.

So using of information is connected with the concept of "knowledge" in a certain subject area.

The subject area - part of the real world, which is considered within a given context. By the context we can be understood, for example, field of study or area that is the subject of some activities. In our case it is a breech block of the artillery. The conceptual model of subject area (CMSA) - a set of concepts, terms and relations between them that correspond to meaning from the real world.

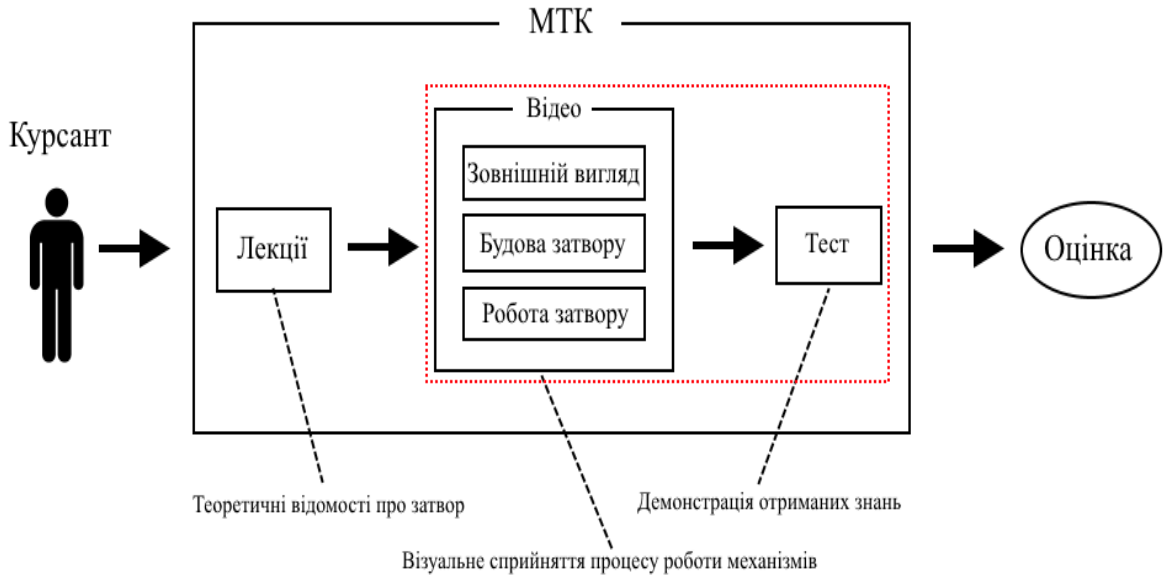


Figure 1. Scheme information technology

II. THE CONCEPTUAL MODEL

The conceptual model of subject area have two components: conceptual and substantive. The conceptual part defines the concepts, terminology of the domain and the relationships between them.

Any relationship should also be a concept, which can be in relations with other concepts. The conceptual model of subject area is a basic element in working with the information, as it defines the "skeleton" on which a substantive component is attached.

Information component of the content model domain correspond to the real domain objects. The relationships described by the conceptual terms of (CMSA) can be divided into two types: substantive and conceptual.

Substantive define the relations of one information element to another, and the conceptual - the relationship of element to the concept with the conceptual (CMSA).

Despite the fact that the graph structure of (CMSA) limits us to using only binary relations, it has no significant effect on the severity of the simulation, because, obviously, any n-pairwise relation can be represented in the form of several binary relations.

The conceptual domain model can be represented in a graph M_d:

$$M_d = < W_d, L_d > \quad (1)$$

where: W_d is the set of vertices, corresponding to the terms of the subject area,

L_d is the set of arcs that represent relationships between the concepts.

Developed graph model depicted in figure 2.

Graph (1), and model view M_d can be further expanded to multiple interpretations of concepts. Obtained in this case structure is actually a special case of the model of the ontology:

$$ONT = < W_d, L_d, F > \quad (2)$$

where: F - final set of interpretation functions defined on the concepts and/or relations of ontology.

In computer science, it is an accepted working definition, according to which ontology is a database of data/knowledge that describes facts that are assumed to be always true within a particular subject area and the particular community of users.

In the private case we have a taxonomy that is defined as:

$$ONT = < W_d, \{is\_a\}, \{\ \} > \quad (3)$$

Taxonomy structure means a hierarchical system of concepts, connected by relations of the form $\{is\_a\}$ "to be part of the class." This type of relations allows to organize the structure of concepts of the ontology in a tree view, that is makes it identical to model (1).

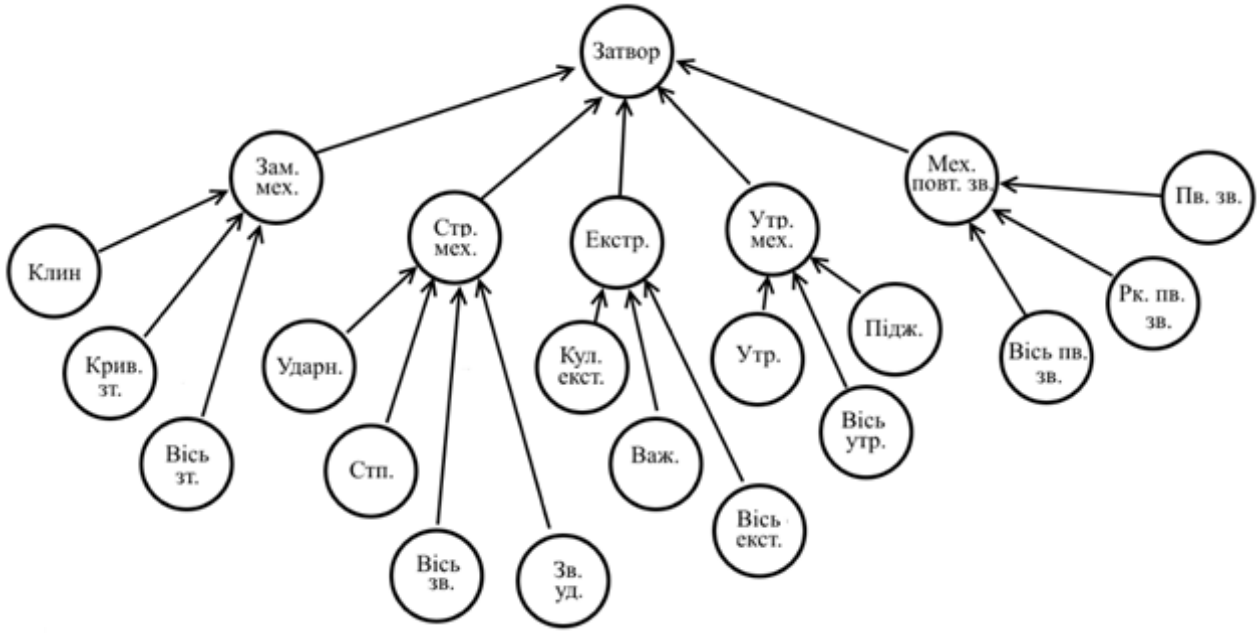


Figure 2. Graph model of shutter

The conceptual model species the M_d, developed with the establishment of a virtual simulator shutter, which includes the 7 mechanisms of basic level: locking, firing, extraction, re-construction, preventive, holding, semi-automatic.

The vertices of the first level, form a number W_{d1}, define the basic concepts of the subject area. Each concept with W_{d1}, is the root vertex to some fragment of the subject knowledge that defines the lower level.

However, the formation of information models (In_Mi) on the basis of the establishment of the equivalence relations of type $\forall_i\left(In\_M_i \leftrightarrow W_{d^1}(i)\right)$ is not appropriate.

Experience of creation of applied intelligent automated systems in the field of research, diagnosis, design and training, shows that development of software based on the mono-alternative set of information models leads to loss of functionality of training tools.

Depending on the purposes and methods of processing information to key concepts (W_{d^1}) it is necessary to develop several different models of representation.

The model reflects the basic concepts of the subject area and existing between generic-specific (class-subclass) relationships L_d. Each concept or conjunction can interpret some fact, define the object or abstract entity. When expanding many L_d by incorporating the set $\{ot_i\}$ and functional relations, causal relations, it is possible to display more complex utterances and relationships between objects.

CONCLUSION

Most serious training systems are complex software and hardware systems. Such embodiment maximizes training. However, such systems require significant financial investments and difficult to maintain. The disadvantages of such simulators is that they are highly specialized for a specific task, and almost no opportunity to change goals.

In some areas of computer simulators severely restricted and permissible only in the initial stages of training.

It is therefore necessary to create virtual or multimedia simulators that perform the task. The advantages of this simulator are a low cost, compactness, possibility of location in almost any room.

REFERENCES:

[5] M.A. Belov, *Principles of Designing the Virtual laboratory of computer technology based on cloud technologies [*Modern problems and solutions WAYS s in science, transport, Production and Education - 2010"]. 2010.

[6] B.V. Palyuh, *E-Learning in engineering education* / Quality Education. №10. 2012. S.34-37.

[7] Y.P. Norenkov, *Information Technology in Education* / Eds. M., Bauman , 2004. 352 p.

[8] N.N. Filatova, N.I. Vavilova, "The trainer-simulator's models of the world on the basis of the plurality of the figurative representations" // International Journal "Information Theories and Applications" - 2001, volume 8, number 4, pp. 176-184.

SESSION 4
ENGINEERING OF INFORMATION SYSTEMS

Comparison of the responsiveness of ultrasonic oscillating temperature sensors (UOTSes) and conventional sensors to temperature inflection points

A.Hashmi, A.Kalashnikov
Sheffield Hallam University, UK, a.kalashnikov@shu.ac.uk, http://tinyurl.com/lokvnbt

***Abstract* – Ultrasonic oscillating temperature sensors (UOTSes), in distinction to conventional temperature sensors, feature almost negligible settling time. This property can be useful for detecting malfunctions, failures and misuses of heat exchangers. However, most exchangers handle substantial thermal masses, which obscure the detection of any temperature changes.**

We compared the responsiveness of conventional DS18B20 sensors and an UOTS to the change in the temperature gradient of over 3.5 kg of water using a posteriori records. Temperature inflection points were estimated by extending the curves for separate distinct heating and cooling intervals that fit best and finding their interception. For the UOTS, the interception occurred about 100 seconds sooner, making it a potential candidate for detecting heat exchangers' irregularities.

Keywords – temperature sensing, ultrasonic oscillating temperature sensors, detecting temperature inflection points

I. INTRODUCTION

Sensing temperature is essential for the appropriate control of industrial processes in many industries (*e.g.*, in the food and petrochemical industries) and domestic appliances (*e.g.*, domestic heaters and refrigerators). The world market of temperature sensors was estimated to be worth over US $5 billion in 2016 [1]. Conventional sensors consist of an encased sensing element that needs to be brought to thermal equilibrium with the environment before taking any measurements. The typical settling times of these sensors are in the range of several seconds, which introduces unwelcome lag when monitoring processes of interest. By their very operating principle, conventional sensors sense temperature at a single point only. For this reason, many sensors need to be procured, installed, wired, serviced, and interrogated if the average temperature in a process vessel or industrial freezer is to be controlled or maintained.

In contrast, UOTSes sense temperatures based on the fact that the velocity of ultrasound depends upon the temperature in the medium of interest. This velocity ranges from hundreds of meters per second for gases to several kilometres per second for liquids and solids, making it possible to interrogate substantial ultrasound pathways at once with no settling time at all.

We are developing UOTSes that can be realised at a cost commensurate with the cost of conventional sensors (Table I, [2]).

TABLE I
PREVIOUS UOTS DEVELOPMENT

Refe-rence	UOTS center fre-quency	Approx i-mate sensiti-vity	Length of the path-way	Comments
[2]	330 kHz	280 Hz/K	0.03 m	Consistency of UOTS output frequencies vs. temperature at decreasing temperatures was reported
[3]	25 kHz	40 Hz/K	0.19 m	Different start up frequencies from the same UOTS in different experiments were observed
[4]	29 kHz	Tilt sensor	0.05 m	Reliable way to measure UOTS output frequency with any required resolution was presented
[5]	22 kHz	50 Hz/K	0.1 m	Implementation options for the electronic driver (including PSoC1*) were discussed
[6]	25 kHz	25 Hz/K	0.1 m	Comparison of ultrasonic thermometer architectures was conducted
[7]	46 kHz	60 Hz/K	0.1 m	Use of a UOTS for overnight measurements and observed hysteresis were reported
[8]	25 kHz	20 Hz/K	0.1 m	Simultaneous use of two UOTS for the same process, modular design of the electronic driver, and thermal hysteresis for the recorded data were discussed
[9]	27 kHz	30 Hz/K	0.1 m	Differential temperature measurement using two UOTS was reported

*PSoC1 refers to the programmable systems on chip series 1 device, which is a highly versatile electronic part manufactured by Cypress Semiconductor.

UOTSes consist of a pair of ultrasonic transducers placed inside a positive feedback loop (Fig.1).

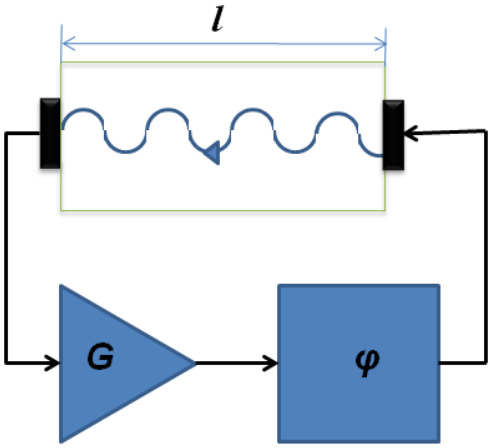


Fig.1. Block diagram of an UOTS

Practical temperature control of most heat exchangers is complicated by their substantial thermal mass. For this reason, a heater, after being switched off, will continue to heat the reservoir for some time until it reaches thermal equilibrium with the reservoir. Detecting the heater's failure is only possible after some time lag that might depend on the operating principle of the temperature sensors used.

In this paper, we compare the responsiveness of conventional DS18B20 sensors to a temperature inflection point to the responsiveness of an UOTS, assessed from *a posteriori* experimental data.

II. Experimental Setup and Procedure

The experiment was conducted by heating/cooling a plastic cylinder containing over 3.5 kg of water through air convection. This cylinder was equipped with six DS18B20 and one UOTS. First, the cylinder was placed into a thermal chamber with a heater that was turned on for a period of time (heating stage) and then was switched off (cooling stage). During the heating stage, the thermal chamber gained the same amount of energy over the same period of time, which should be possible to approximate by determining the linear dependence of the temperature versus time. At the second stage of the experiment, natural cooling should have obeyed Newton's law of cooling [11], and the chamber's temperature would be expected to exponentially decrease over time. After recording readings from all the sensors over the course of the experiment, we approximated these sensor readings separately for the heating and cooling stages by best fit curves. The interceptions of the best fit curves gave estimates for the temperature inflection points.

Fig. 2, left, presents the average temperature readings from the conventional sensors. The recorded output UOTS frequencies are plotted in Fig. 3, left. These frequencies were further processed to eliminate the intermittent deviations of the output frequency from the smooth, long-term curve as shown in the centre of Fig. 2.

III. Estimation of the Temperature Inflection Points Form the Recorded Data

Suitable heating and cooling intervals, where the readings of all the sensors changed steadily, were selected first. Because ultrasonic and temperature data were collected at different sampling rates, these intervals differ slightly as shown in the centre graphs of Fig.2 and Fig.3. These differences should not have notably affected the curves, as they are fitted over substantial time intervals. The parameters of the best fit curves for both the heating and cooling stages were independently calculated using the function **fit** of MATLAB.

The best fit interpolation curves for the average temperature were found to be linear (heating) and exponential (cooling) ones as expected. When plotted, these curves matched the corresponding average temperature curve segments well (Fig. 2, centre), and their extensions intersected at around t=2955 s (Fig. 2, right).

Approximating the UOTS output frequency curves required some additional considerations. An oscillator sustains steady state oscillations if it satisfies the Barkhausen criterion [12]: an UOTS must compensate for all the signal loop losses and provide zero end-to-end phase shift. The latter condition depends on the ultrasound TOF between the transducers (hence on the ultrasound velocity which depends on the temperature), making the UOTS feasible. Fig. 1 presents a block diagram for a UOTS, showing both the ultrasonic and electronic devices and the signal pathways relevant to the sensor's operation. The phase condition can be written as follows:

$$\frac{l}{\lambda} + , \qquad (1)$$

where λ is the ultrasound wavelength, , c is the ultrasound velocity, l is the distance between the ultrasonic transducers, f is the UOTS output frequency and is the overall phase shift in the positive feedback loop.

Assuming that, as the first order approximation, G, l, and are all independent of the ambient temperature and the UOTS output frequency, the wavelength of the standing wave between the transducers should remain unchanged, despite any changes in the ultrasound velocity caused by the temperature. Hence, if the velocity changed by due to some temperature change, the corresponding change in the UOTS frequency should satisfy the following relations:

$$\lambda = const \;\Rightarrow\; \frac{c+\Delta c}{f+\Delta f} = \frac{c}{f} \;\Rightarrow\; \frac{\Delta c}{c} = \frac{\Delta f}{f}$$

$$\Rightarrow \quad \Delta f = \frac{\Delta c}{c} f \qquad (2)$$

The derivative can be estimated by using either seminal Del Grosso and Mader fifth order polynomial approximation [13] or the considerably lighter limited applicability second order approximation proposed by Lubers and Graaff [14],

$$c(T) = 1404.3 + 4.7\,T^2, \qquad (3)$$

where T is expressed in degrees Celsius.

The second order approximation seems appropriate here because of the temperature range of interest and the consideration of the electronic and mechanical components of the UOTS as being ideal. It follows that

$$\frac{d}{d'}$$

$$\Delta f = f \frac{4.7 - 0.08\,T}{1404.3 + 4.7\,T - 0.04\,T^2} \Delta T \Rightarrow$$

$$\frac{\Delta T}{\Delta f} = \frac{1404.3}{f(} \qquad (4)$$

Equation (4) stipulates that the UOTS output frequency depends on the temperature in a non-linear fashion, and the UOTS sensitivity depends on the ambient temperature and nominal UOTS operating frequency. Both of these conclusions were observed experimentally. The experimental UOTS output frequencies, recorded at the heating stage, were best approximated by a second order polynomial, while the output frequencies at the cooling states were best fit with an exponential curve. These curves intersected at around t = 2852 s (Fig. 3, right). Therefore, the UOTS data allowed for detecting the inflection point at approximately 100 s faster than the conventional sensor data *a posteriori*.

IV. Summary and Conclusions

Detecting heat exchanger irregularities can be complicated by the substantial thermal masses involved in the exchange process. We estimated the location of the detected temperature inflection points using *a posteriori* data recorded by both conventional temperature sensors and an UOTS. Suitable curves were fitted to distinct intervals of data related to the heating and cooling stages, and the curves' interceptions were used to locate the temperature inflections. The temperature inflection point, estimated from the conventional sensors' data, lagged by around 100 seconds compared to the point estimated from the UOTS data. This result supports the case for the application of UOTS for failure/malfunction/misuse detection in heat exchangers.

References

[1] Temperature Sensors Market by Type (Bimetallic, Thermistor, IC, RTD, Thermocouple, IR, Fiber Optics, Others (MEMS, USB, Wi-Fi, ZigBee, Bluetooth, RFID)), End User (Process Industries & Discrete Industries), & Geography - Global Forecast to 2022, summary available online on tinyurl.com/jrpj9dr, accessed Apr 2016.

[2] S. Alzebda and A. N. Kalashnikov, "Ultrasonic sensing of temperature of liquids using inexpensive narrowband piezoelectric transducers," in *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 57, no. 12, pp. 2704-2711, December 2010. doi: 10.1109/TUFFC.2010.1744.

[3] A. Afaneh, S. Alzebda, V. Ivchenko, and A. N. Kalashnikov, "Ultrasonic Measurements of Temperature in Aqueous Solutions: Why and How," in *Physics Research International*, vol. 2011, Article ID 156396, 10 pages, 2011. doi:10.1155/2011/156396.

[4] P. Popejoy, S. Alzebda, O. Sonbul and A. N. Kalashnikov, "Linear angle measurement using continuous wave ultrasonic oscillator," *Instrumentation and Measurement Technology Conference (I2MTC), 2012 IEEE International*, Graz, 2012, pp. 733-736. doi: 10.1109/I2MTC.2012.6229191.

[5] P. Popejoy *et al*., "Comparison of implementations of driving electronics for ultrasonic oscillating sensors," *2012 IEEE International Ultrasonics Symposium*, Dresden, 2012, pp. 2427-2430. doi: 10.1109/ULTSYM.2012.0607.

[6] A.Hashmi, M.Malakoutikhah, R.A.Light, A.N.Kalashnikov, " Embedded Supervisory Control and Output Reporting for the Oscillating Ultrasonic Temperature Sensors," in *Proc. 4th Computer Science On-line Conference 2015 (CSOC2015), Vol 2: Intelligent Systems in Cybernetics and Automation Theory*, Springer, 2015, pp.139-150. doi: 10.1007/978-3-319-18503-3_14.

[7] A.Hashmi, A.N.Kalashnikov and R.A.Light, "Inexpensive differential ultrasonic calorimeter for accurate measurement of heat loss," in *Proceedings of the Eighth Saudi Students Conference in the UK*, World Scientific, 2015, pp.525-536.

[8] A. Hashmi, A. N. Kalashnikov and R. A. Light, "Experimental evaluation of ultrasonic oscillating temperature sensors (UOTS) under cyclically changing temperatures," *Ultrasonics Symposium (IUS), 2015 IEEE International*, Taipei, 2015, pp. 1-4. doi: 10.1109/ULTSYM.2015.0249.

[9] A.Hashmi, R.A.Light, M.Malakoutikhah and A.N.Kalashnikov, "Accurate heat loss evaluation of water-cooled electric motors using a differential ultrasonic calorimeter," A. Hashmi, R. A. Light, A. N. Kalashnikov and M. Malakoutikhah, "Accurate heat loss evaluation of water-cooled electric motors using a differential ultrasonic calorimeter," *2016 International Conference for Students on Applied Engineering (ICSAE)*, Newcastle upon Tyne, 2016, pp. 42-47. doi: 10.1109/ICSAE.2016.7810158.

[10] A.Hashmi, M.Malakoutikhah and A.N.Kalashnikov, "Measuring thermal losses of water cooled motors using oscillating ultrasonic temperature sensors", in *Proc. II Int. Conf. Advanced Information Systems and Technol.*, Sumy, Ukraine, 2014, pp.130-132.

[11] Newton's law of cooling, Wikipedia, available online on tinyurl.com/z4lmetp, accessed Sept 2016.

[12] Barkhausen stability criterion, Wikipedia, available online on tinyurl.com/c87ugzz, accessed Sept 2016.

[13] V. A. Del Grosso and C. W. Mader, "Speed of sound in pure water," in *J. Acoust. Soc. Am.*, vol.52, no. 5, pp. 1442–1446, 1972. doi: 10.1121/1.1913258.

[14] J.Lubbers and R.Graaff, "A simple and accurate formula for sound velocity in water", in *Ultrasound Med. Biol.*, 1998, vol.24(7), pp.1065-1068. doi: 10.1016/S0301-5629(98)00091-X.

Fig.2. Average temperatures over the course of the experiment (left), with the heating and cooling stages shown (centre) and the estimate of the temperature inflection point (right) (REF=25.7°C)

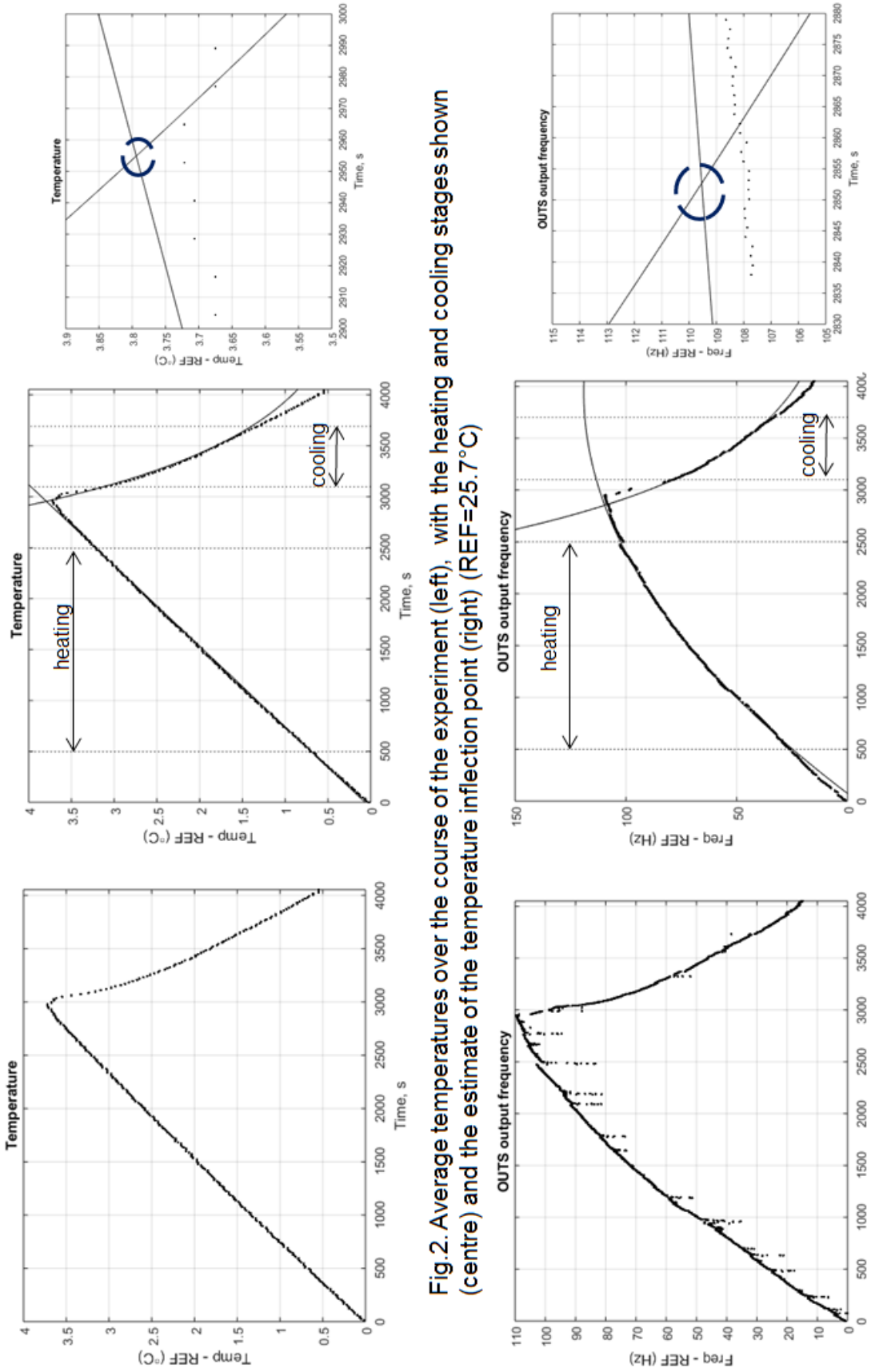


Fig.3. Recorded raw UOTS output frequencies (left), processed frequencies with the heating and cooling stages shown (centre) and the estimate of the temperature inflection point (right) (REF=27168 Hz)

Enhancement of low quality degraded video using haar wavelet decomposition technique

[1]Ms.Prajkta Gupta M.E(Scholar), Prof.V.B.Gadicha Asstt.Professor
PR Patil College of Engineering & Technology (PRPCET), Amravati, Maharashtra, India
Prajktaguptapaper2017@gmail.com , vbgadicha@gmail.com

***Abstract* – The point recommends the general system for video improvement, object location and following continue at the same time in an aggressive and helpful way .Video is a gathering of successive pictures with a consistent time interim. So video can give more data about the article when situations are changing regarding time. Thusly, physically taking care of recordings are entirely unimaginable. So there is need of a computerized gadget to handle these recordings. Numerous calculations and innovation have been created to robotize video improvement and checking the article in a video document. Video improvement with item identification and following is a one of the testing errand in PC vision. Additionally, following of an article essentially includes two going before steps object identification and item representation. Object discovery is performed to check presence of items in video and to correctly find that article .Object following is a procedure of dividing a district of enthusiasm from a video scene and monitoring its movement, position and impediment. The following is performed by observing articles' spatial and transient changes amid a video grouping, including its nearness, position, size, shape, and so forth. Video improvement with article following is utilized as a part of a few applications, for example, video reconnaissance, robot vision, movement observing.**

Keywords – Haar Wavelet , Video Enhancemnet,Frame Rate, Mean Intensity.

I. INTRODUCTION:

Video surveillance is an active research topic in computer vision that tries to detect, recognize and track objects over a sequence of images and it also makes an attempt to understand and describe object behavior by replacing the aging old traditional method of monitoring cameras by human operators. Object detection and tracking are important and challenging tasks in many computer vision applications such as surveillance, vehicle navigation and autonomous robot navigation. Object detection involves locating objects in the frame of a video sequence. Every tracking method requires an object detection mechanism either in every frame or when the object first appears in the video. Object tracking is the process of locating an object or multiple objects over time using a camera. The high powered computers, the availability of high quality and inexpensive video cameras and the increasing need for automated video analysis has generated a great deal of interest in object tracking algorithms. There are three key steps in video analysis,

- detection interesting moving objects
- tracking of such objects from each and every frame to frame

analysis of object tracks to recognize.

I.1 OBJECTIVES

- Enhancement of low degraded video to realize the great quality video and higher frame quality.
- The separation of Associate in Nursing audio and frames in uncompressed low vision video to reinforce inferiority video and perform object detection.
- To Improve the speed and accuracy of technique used for police work target image at intervals less time.
- To increased frames of pictures that works well below image blur, camera motion, modification of cause, illumination, and scale conditions
- To propose increased object following techniques supported following rule used template matching methodology.
- To find target object and match each frames in video
- To bring home the bacon Noise Free Video.

I.2 SCOPE

Many algorithms and technology have been developed to automate monitoring the object in a video file. Object detection and tracking is a one of the challenging task in computer vision. Also, tracking of an object mainly involves two preceding steps object detection and object representation. Object detection is performed to check existence of objects in video and to precisely locate that object .Object tracking is a process of segmenting a region of interest from a video scene and keeping track of its motion, position and occlusion. The tracking is performed by monitoring objects' spatial and temporal changes during a video sequence, including its presence, position, size, shape, etc. Object tracking is used in several applications such as video surveillance, robot vision, traffic monitoring.

II. LITERATURE SURVEY

The research conducted so far for object detection and tracking objects in video surveillance system are discussed in this chapter. The set of challenges outlined above span several domains of research and the majority of relevant work will be reviewed in the upcoming chapters. In this section, only the

representative video surveillance systems are discussed for better understanding of the fundamental concept. Tracking is the process of object of interest within a sequence of frames, from its first appearance to its last. The type of object and its description within the system depends on the application. During the time that it is present in the scene it may be occluded by other objects of interest or fixed obstacles within the scene. A tracking system should be able to predict the position of any occluded objects. Object tracking systems are typically geared towards surveillance application where it is desired to monitor people or vehicles moving about an area.

[M. kim, 2014] propose a novel framework for enhancement of very low-light video. For noise reduction, motion adaptive temporal filtering based on the Kalman structured updating is presented. Dynamic range of denoised video is increased by adaptive adjustment of RGB histograms. The proposed method exploits color filter array (CFA) raw data for achieving low memory consumption. The adaptive temporal filter based on the Kalman filter and adopted the NLM denoising for further smoothing. Histogram adjustment using the gamma transform and the adaptive clipping threshold is also presented to increase the dynamic range of the low-light vid234eo.

[W. Zhong, 2014] proposed method based on a sparse collaborative model that exploits both holistic templates and local representations to account for drastic appearance changes. Develop a sparse discriminative classifier (SDC) and sparse generative model (SGM) for object tracking. In the SDC module, we present a classifier that separates the foreground object from the background based on holistic templates. It plays a critical role in numerous vision applications such as motion analysis, activity recognition, visual surveillance and intelligent user interfaces. Local representations are adopted to form a robust histogram that considers the spatial information among local patches with an occlusion handling module, which enables our tracker to better handle heavy occlusions.

[N. Kumar, 2015] proposed Mean and Median image filtering algorithms are compared based on their ability to reconstruct noise affected images. The purpose of these algorithms is to remove noise from a signal that might occur through the transmission of an image. These algorithms can be applied to one-dimensional as well as two-dimensional signals. A new framework for removing impulse noise from images is presented in which the nature of the filtering operation is conditioned on a state variable defined as the output of a classifier that operates on the differences between the input pixel and the remaining rank-ordered pixels in a sliding window. In this comparison of noise removal filters, the experiment has been conducted for different images and at various noise levels, and is seen that Median filters performed the best overall noise compositions tested by providing minimum MSE.

2.1 PROBLEM FORMULATION

In this thesis our aim is to improve the performance of object detection and object tracking for low vision video. The detection and recognition of objects proceed simultaneously with image segmentation in a competitive and cooperative manner. Problem occurred in variety of dynamic environments, it has a strong adaptability, but it is generally difficult to obtain complete outline of moving object, responsible to appear the empty phenomenon, as a result the detection of moving object is not accurate. So need to improve the technique.

To achieve this, the following specific objectives

To propose object detection techniques that to effectively identify objects of interest area in video sequence.

To propose frames enhancement techniques that improve the quality of degraded frames by using enhance frame, filter object and spatial noise reduction algorithms.

To propose enhanced object tracking techniques based on tracking algorithm used template matching method.

III. PROPOSED SYSTEM

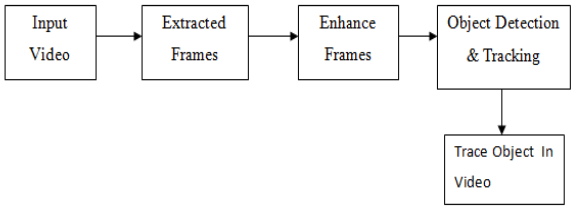


Fig 3.1: Architecture of Proposed Method

3.1 DATA FLOW DIAGRAM

Algorithm 3.1.2

Input : - Video

Output: - 1. Show trace object into frames
2. Show trace object into video
3.Show the difference between original frame and enhance frame (Result analysis).

Step 1: Take sample video as input -Read the sample video file.

Step 2: Extract frames and sound separately in video file.

Step 3: Decomposed Frames into Haar Wavelet Transform.

Step 4: Check the video quality

If (Low video quality: need to enhance)

Else (no need to enhance)

Step 5: Enhance frame (enhance the quality of frame used filter object And spatial noise reduction method)

Step 6: Load enhance frame

Step 7: Get track object into enhance frame

Step 8: Check tracking object into all enhance frame (with strong feature)

Step 9: Store and display the trace object with square box into video.

Step 10: Show the difference between the original frame and enhance frame.

Algorithm 3.1..3: Quality Decision

Step 1: Start
Step 2: Read Extract frames.
Step 3: Check the video quality
For i=1: length(Efi)
Read EFi
If M.I.(Efi) > threshold
Add to Quality frame EFi
Else
Add EFi to Degraded frame
End
End
Step 4: Stop

In above algorithm: M.I- Mean Intensity EFi- Extract frame for 1 to length. Here threshold if fixed.

Algorithm 3.1..4: Enhance frames

Step 1: Start
Step 2: Read Degraded frames
Step 3: For i=1: length(Dfi)
Read DFi
If((MI<=0.1)&&(MI>0.0))
Width=0.7;
End
If((MI>0.1)&&(MI<=0.2))
Width=0.75;
End
If((MI>0.2)&&(MI<=0.3))
Width=0.8;
End
If((MI>0.3)&&(MI<=0.4))
Width=0.85;
End
If((MI>0.4)&&(MI<=0.5))
Width=0.9;
End
Step 4: Save DFi
End
Step 5: Stop

In that compare mean intensity of degraded frame and set new enhance factor of degraded frame. Improve the quality of degraded frames.

Algorithm 3: Filter object algorithm

Step 1:Start
Step 2:Read all enhance frame from enhance algorithm
Step 3: For i=1: length(Efi)
Read EFi
EFi = Histogram Equalize (EFi)
Save EFi
End
Step 4: Stop

IV. RESULT ANALYSIS

Table 4.2 Result Analysis

Sr No	Methods	Total Frames	Enhance Frames	Quality
1	Proposed	50	48	96
2	Histogram Equalization	50	35	70
3	Weber's Law	50	32	64
4	Spatial Noise	50	43	86
5	Constrast Enhancement	50	23	46
6	Invert Method	50	28	56

Table 4.3 Object Detection Analysis

Sr.No	Object Present	Object detected	True Detection Rate(%)	False Detection Rate	Error Rate
1	10	8	80	3	20
2	15	14	93.33	5	6.67
3	20	18	90	2	10
4	25	22	88	2	12
5	30	27	90	1	10

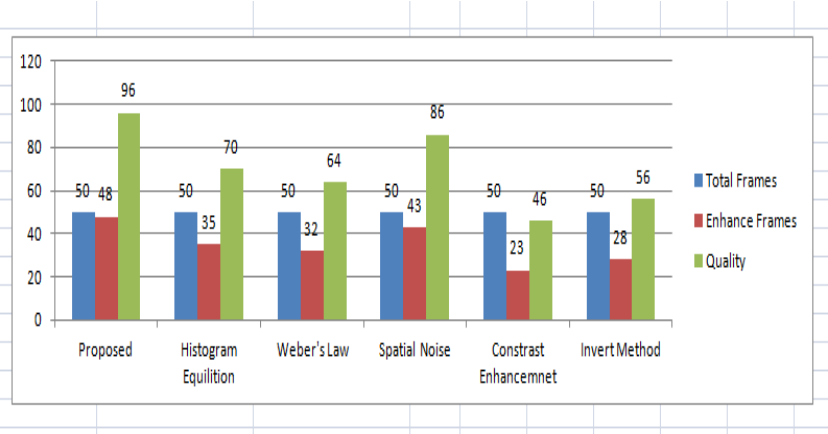


Figure 4.1 – Graph Analysis for table 4.2

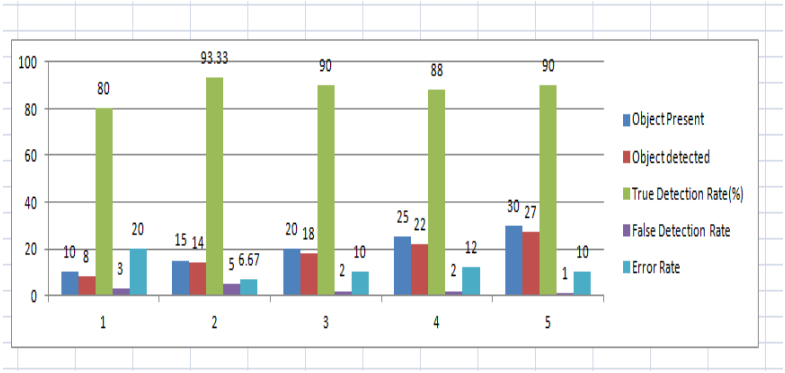


Figure 4.2 – Graph Analysis for table 4.3

CONCLUSIONS

From above analysis, it is observed that, proposed methodology gives better result even for low light video too. That helps in intruder detection, security application.

REFERENCES

[1] A. Yilmaz, a., javed, o., and shah, m. 2006. Object tracking: a survey. Acm comput. Surv. 38, 4, article 13.

[2] B. Karasulu 2010, "review and evaluation of well-known methods for moving object detection and tracking in videos", journal of aeronautics and space technologies

[3] Kaiqi Huang, Liangsheng Wang , Tieniu Tan & Steve Maybank 2007, "A real-time object detecting and tracking system for outdoor night Surveillance", Published by Elsevier Ltd on behalf of Pattern Recognition Society.

[4] Abhishek Kumar Chauhan, Prashant Krishan, "Moving Object Tracking Using Gaussian Mixture Model And Optical Flow", International Journal of Advanced Research in Computer Science and Software Engineering, April 2013

[5] N. Prabhakar, v. Vaithiyanathan, a.sharma, a. Singh and p singhal 2012," object tracking using frame differencing and template matching", research journal of applied sciences, engineering and technology.

[6] T. Waykole and y. Jain 2013," detecting and tracking of moving objects from video", international journal of computer applications (0975 – 8887).

[7] K. Ahuja and p. Tuli 2013," object recognition by template matching using correlations and phase angle method", international journal of advanced research in computer and communication engineering

[8] S. Shantaiya, k. Verma and k. Mehta 2013," study and analysis of methods of object Detection in video", international journal of soft computing and engineering (ijsce)

[9] Kavita Ahuja , Preeti Tuli ," Object Recognition by Template Matching Using Correlations and Phase Angle Method", International Journal of Advanced Research in Computer and Communication Engineering

[10] G.Sindhuja and Dr. R.Devi S.M. 2015," A Survey on Detection and Tracking of Objects in Video Sequence", International Journal of Engineering Research and General Science.

[11] W. Zhong , huchuan lu and m. Yang 2014," robust object tracking via sparse Collaborative appearance model", ieee transactions on image processing.

[12] A brief study of various noise model and filtering techniques, Versha Rani et al, Journal of Global Research in Computer Science, 4 (4), April 2013, 166-171

Creating Highly Available Distributed File System for Maui Family Job Schedulers

Andrii Onishchuk
National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute"

This article describes a way to implement a distributed file system for MAUI job scheduler, which solves the problems of low scalability and unreliability of data storage, as well as a problem of problem of data inaccessibility due to failures in software or hardware. The architecture which is suitable for MAUI GRID systems is suggested.

Keywords **– distributed file system, MAUI task scheduler, file system scalability, high reliability of file systems, name node, datanode.**

I. INTRODUCTION

In recent years a rapid increase in popularity of distributed systems has been seen. The reason this is in their greater reliability, scalability and power. That's why a need for simple and convenient software that simplifies the user experience for such a systems has drastically increased.

Currently to run MAUI task user of GRID system must manually copy the files to each node in the system, which increases the likelihood of errors. Also this approach has a number of other disadvantages compared to using the concept of a distributed file system for organizing the files workflow. The advantages of using this concept are scalable and reliable data storage, data accessibility and low cost of equipment for storing files.

II. ARCHITECTURE OF DISTRIBUTED FILE SYSTEM FOR MAUI

Distributed File System (DFS) is a file system where the file pieces (blocks) are stored on a bunch of computers connected with high-bandwidth network [1]. The system which is described in this article is a subtype of DFS, which is used for Maui Scheduler.

The system has two main types of nodes: namenode and datanode. And two types of ancillary nodes: journalnode and standby namenode. Figure 1 shows the architecture, placement and interaction of these tytpes nodes. Namenode is a master node in DFS. There is only one active namenode in the DFS. It stores metadata of files, as well as information about where data are stored in the cluster file. Metadata examples are file names, their types, permissions, data about blocks location within the network. Namenode does not store any file blocks. This is done in order to reduce the load of it. In most file operations first, and sometimes the only call is made to namenode. The exception is a write operation to a file, which requires coordination between all types of file system nodes.

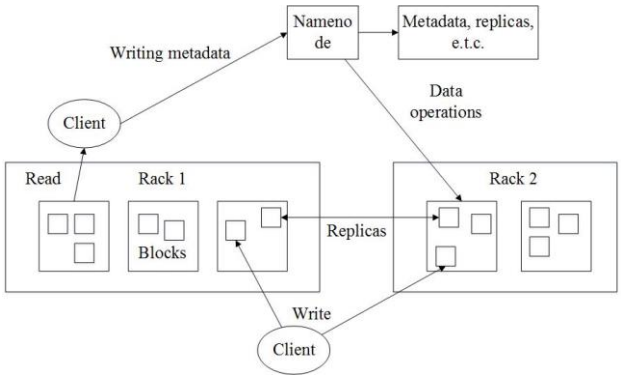


Figure 1. Architecture of RFS

MAUI jobs communicate with namenamenode every time they want to find a file, or add / copy / move it.

Datanode – is a slave node (master/slave achritecutre), the main purpose of which is to preserve data blocks. To take advantage of the RFS one should have more than a single datanode in a GRID system. Each node is aware of the blocks on it. The blocks are accessible namenode tells the user an exact location of them of datanode. In addition, datanode can replicate file blocks to improve system reliability. Also due to replication, this type of nodes usually does not require administrators to install RAID drives.

How does the namenode choose which datanodes to store replicas on? There's a tradeoff between reliability and write bandwidth and read bandwidth here. For example, placing all replicas on a single node incurs the lowest write bandwidth penalty since the replication pipeline runs on a single node, but this offers no real redundancy (if the node fails, the data for that block is lost). Also, the read bandwidth is high for off-rack reads. At the other extreme, placing replicas in different data centers may maximize redundancy, but at the cost of bandwidth. Even in the same data center (which is what all Maui GRID systems to date have run in), there are a variety of placement strategies. Indeed, MAUI changed its placement strategy in release 0.17.0 to one that helps keep a fairly even distribution of blocks across the cluster. And from 0.21.0, block placement policies are pluggable.

MAUI's default strategy is to place the first replica on the same node as the client (for clients running outside the cluster, a node is chosen at random, although the system tries not to pick nodes that are too full or too busy). The second replica is placed on a different rack from the first (off-rack), chosen at random. The third

replica is placed on the same rack as the second, but on a different node chosen at random. Further replicas are placed on random nodes on the cluster, although the system tries to avoid placing too many replicas on the same rack.

Once the replica locations have been chosen, a pipeline is built, taking network topology into account. For a replication factor of 3, the pipeline might look like Overall, this strategy gives a good balance among reliability (blocks are stored on two racks), write bandwidth (writes only have to traverse a single network switch), read performance (there's a choice of two racks to read from), and block distribution acrossthe cluster (clients only write a single block on the local rack).

For fast file blocks access, namenode caches often used blocks. Therefore, increasing the number of nodes on the RAM data although it may result in a slightly faster performance, but not critical.

Another part of the file system is an RFS client. RFS client is a software library that allows you to work with the file system using simple unix-like commands. Allowed commands are: ls, rm, mkdir, touch, some other RFS specific commands are: copyFromLocal, copyToLocal. These commands allow you to copy files or folders from the local file system to RFS and vice versa. Another RFS client is MAUI RFS client integrated with MAUI API interface. It is used to do file operations through built MAUI interface [2].

Let's take a look at the example of the interaction of the nodes in case of basic file operations. The most complex operation in terms of nodes interaction is the RFS file write operation (WRITE). It consists of four steps described below:

1. Call to namenode for a list of datanodes which should receive file blocks.
2. Uploading of file blocks to the given nodes.
3. Replication of the received blocks by datanodes.
4. Sending the information about the file blocks location to the namenode.

Operation of changing file attributes (CHANGEATTR) such as permissions, name, location is done with one call to namenode. The same applies to the file removal operation (RMFILE), creating a folder (MKDIR) or empty file (TOUCH). File copy operation is implemented through a series of calls from namenode to datanodes.

Let's define RFS file block concept in greater details. RFS blocks are large compared to disk blocks, and the reason is to minimize the cost of seeks. By making a block large enough, the time to transfer the data from the disk can be significantly longer than the time to seek to the start of the block. Thus the time to transfer a large file made of multiple blocks operates at the disk transfer rate.

A quick calculation shows that if the seek time is around 10 ms and the transfer rate is 100 MB/s, to make the seek time 1% of the transfer time, we need to make the block size around 100 MB. The default is actually 64 MB, although many HDFS installations use 128 MB blocks. This figure will continue to be revised upward as transfer speeds grow with new generations of disk drives.

Let's describe the permission model in our DFS. It has a permissions model for files and directories that is much like the POSIX model. There are three types of permission: the read permission (r), the write permission (w), and the execute permission (x). The read permission is required to read files or list the contents of a directory. The write permission is required to write a file or, for a directory, to create or delete files or directories in it. The execute permission is ignored for a file because you can't execute a file on DFS (unlike POSIX), and for a directory this permission is required to access its children [3].

Each file and directory has an owner, a group, and a mode. The mode is made up of the permissions for the user who is the owner, the permissions for the users who are members of the group, and the permissions for users who are neither the owners nor members of the group.By default, DFS runs with security disabled, which means that a client's identity is not authenticated. Because clients are remote, it is possible for a client to become an arbitrary user simply by creating an account of that name on the remote system. This is not possible if security is turned on. Either way, it is worthwhile having permissions enabled (as they are by default; see the dfs.permissions.enabled property) to avoid accidental modification or deletion of substantial parts of the filesystem, either by users or by automated tools or programs.

III. Organization for highly reliable distributed file system.

The combination of replicating namenode metadata on multiple filesystems and using the secondary namenode to create checkpoints protects against data loss, but it does not provide high-availability of the filesystem. The namenode is still a single point of failure (SPOF). If it did fail, all clients—including Maui jobs—would be unable to read, write, or list files, because the namenode is the sole repository of the metadata and the file-to-block mapping. In such an event the whole Maui system would effectively be out of service until a new namenode could be brought online.On large clusters with many files and blocks, the time it takes for a namenode to start from cold can be 30 minutes or more.We remedy this situation by adding support for RFS.

In this implementation there is a pair of namenodes in an active-standby configuration. A few architectural changes are needed to allow this to happen:

- The namenodes must use highly-available shared storage to share the edit log. When a standby namenode comes up, it reads up to the end of the shared edit log to synchronize its state with the active namenode, and then continues to read new

entries as they are written by the active namenode.
- Datanodes must send block reports to both namenodes because the block mappings are stored in a namenode's memory, and not on disk.
- Clients must be configured to handle namenode failover, using a mechanism that is transparent to users.
- The secondary namenode's role is subsumed by the standby, which takes periodic checkpoints of the active namenode's namespace.

If the active namenode fails, the standby can take over very quickly (in a few tens of seconds) because it has the latest state available in memory: both the latest edit log entries and an up-to-date block mapping. The actual observed failover time will be longer in practice (around a minute or so), because the system needs to be conservative in deciding that the active namenode has failed. One of the main advantages of a distributed file system is the possibility of highly reliable file storage and access. The problems that might disrupt high reliability access to data or damage the data are [4]:

- hardware or software failure of namenode
- hardware or software failure of datanodes
- unavailability of data due to network problems

Hardware or software failure of datanode is solved using file blocks replication. Replication in this sense is excessive copying of data blocks between datanodes. The number of nodes which store data blocks is called replication factor. By setting up a high replication factor and replication to different network segments we can achieve reliable access to the files in the event failure of multiple nodes at once, or even of the entire network segments. In order to for file system to work when namenode failure occurs, the introduction of two types of auxiliary nodes is required. They are journalnode and stanby namenode.Two separate machines are configured as namenodes. At any given time, exactly one namenode is active and the other is in standby. Active namenode is responsible for all client operations in the cluster, while other one is in standby mode and is not used, however it still retains enough information about the state of the file system to ensure a rapid transition to it, if necessary. For standby namenode to be synchronized with the active node, both namenodes are connected with a group of individual nodes, so-called journalnodes. When any namespace change is performed by active node, it registers record modification to the journal nodes. Standby namenode is capable of reading the log journalnode and constantly monitors changes in it. As a standby namenode sees the changes, it applies them to its own namespace. In case of failure of the active namenode, reserve one, after reading all the logs from the journalnodes declares itself active. This ensures that the namespace is fully synchronized before the transition to another node happens. In order to provide a fast failover, it is also necessary that the standby node have up-to-date information regarding the location of blocks in the cluster. In order to achieve this, the datanodes are configured with the location of both namenodes, and send block location information and heartbeats to both.It is vital for the correct operation of an GRID system that only one of the namenodes is active at any point in time. Otherwise, the namespace state would quickly diverge between the two, risking data loss or other incorrect results. In order to ensure this property and prevent the so-called "split-brain scenario," the journalnodes will only ever allow a single namenode to be a writer at a time. During a failover, the namenode which is to become active will simply take over the role of writing to the journalnodes, which will effectively prevent the other namenode from continuing in the active state, allowing the new active to safely proceed with failover.Inaccessibility due to network problems can be easily eliminated by organizing redundant network topology.The underlying concept behind network redundancy is to provide alternate paths for data to travel along in case a cable is broken or a connector accidentally un-plugged. However, Ethernet as standard cannot have rings or loops in the network as this will cause broadcast storms and can ultimately cause the network to stop working. An Ethernet network cannot have two paths from point A to point B without a mechanism in place to support this type of topology [5]. To achieve redundancy, the network infrastructure (switches) must support redundancy protocols designed to negate the usual problems of putting loops into an Ethernet network, maintaining a default data path and switching to an alternate one when a fault occurs.

CONCLUSIONS

This paper described the architecture of a distributed file system for MAUI, which achieves greater scalability and high reliability and availability of data storage.

REFERENCES:

[1] Ilya Ganelin. Spark: Big Data Cluster Computing in Production, Wiley, March 2016. p.21-32.
[2] Site Maui [Electronic resource] // Access mode: http://www.adaptivecomputing.com/products/open-source/maui/
[3] Neal Kobel. Distributed File Systems: Distributed Computing Architecture, CreateSpace Independent Publishing Platform, December 2016. p.41-67
[4] EC-Council. Computer Forensics: Investigating File and Operating Systems, Wireless Networks, and Storage (CHFI), 2nd Edition (Computer Hacking Forensic Investigator) 2nd Edition, Course Technology, April 2016. p.18-26
5. A. Tanenbaum, Computer Networks 5th By Andrew S. Tanenbaum (International Economy Edition), Prentice Hall, January 2010. p.648-697

The Vth International Conference «Advanced Information Systems and Technologies, AIST 2017»
17-19 May 2017, Sumy, Ukraine

Use of sound for adaptive control of the materials cutting process

R. V. Polishchuk[1], V. V. Nahornyi[1], O. V. Zagovora[2]
[1]Sumy State University, Sumy, Ukraine, vnagornyi1989@gmail.com
[2]GESIS Leibniz Institute for the Social Sciences, Cologne, Germany, Olga.Zagovora@gesis.org

***Abstract* – *A*daptive control of cutting process using as initial data information signals of various physical natures, characterizing indirectly quality of part cutting. The article demonstrates that among a variety of information signals the cutting sound is the most informative. Firstly, it is registered in a contactless manner, making it noise-resistant to spurious signals generated by operating of a metal-cutting tool, and secondly, the cutting sound originating at the point of contact of the tool and the workpiece, has high sensitivity to tool wear and roughness of the machined surface. Further research should focus on the development of processing quality-forecasting methods to promptly change the cutting modes and therefore prolong the period of the defect-free part processing.**

***Keywords* – tool wear; surface roughness; information signal; sound of cutting.**

I. INTRODUCTION

The quality of product manufacturing must be monitored throughout the entire technological chain, including the process of material cutting operations. Thus, the processing quality is understood as compliance of the workpiece geometry and processed surface roughness with the requirements of technical documentation. The compliance is provided by the adaptive control of the processing system operation.

The decisive influence on the quality of controlled parameters is made by the cutting tool wear, which is simultaneously a cause of permanent change in the dynamics of processing equipment. This fact requires implementing of adaptive control with the purpose of varying cutting modes that allow in conditions of ever-changing equipment dynamics to prolong the period of defect-free performance of a predetermined operation of machining materials by cutting.

There are two ways of controlling tool wear-direct and indirect. In industrial practice, the indirect way became widespread. In this regard, the adaptive control of tool wear is indirectly evaluated by measuring information signals having various physical nature and accompanying the cutting process.

Effective performance management is dependent upon the noise resistance of the information signal and its sensitivity to the degree of tool wear and the workpiece surface roughness. These properties of the controlled signal are necessary for the effective management of defect-free workpiece manufacturing, because they eliminate the risk of giving erroneous order.

The article provides a rationale for the selection of suchlike information signal corresponding to the specified requirements.

II. FORMULATION OF THE PROBLEM

The system of adaptive control of the cutting process is based on the measurement of different nature information signals. This signals representing a variety of external factors: mechanical stress, vibration, elastic deformations of the processing system, the electric current, chemical exposure, and the like, which have a decisive influence on the degree of tool wear and, consequently, on the quality of the manufactured product [1].

Therefore, in the process of tool wear such parameters as cutting force [2], the torque [3] and cutting power [4] undergo changes. These parameters are measured by dynamometers. Applying the vibration sensor being disposed, for example, when turning on the tool holder, the tool vibration accompanying inevitably the cutting process is measured [5]

In the course of the wearing tool interaction with a work piece, acoustic emission waves are generated. This signals are recorded by sensors of acoustic emission in the frequency band from 1 kHz to 1 MHz. Acoustic emission is more sensitive than the force factors and vibration to tool wear [6], but at the same time, it makes acoustic emission more sensitive to noise disturbance caused by the influence of the environment and the work of the structural units of the machine.

For this reason, the registration of "integrated settings" is referred to, as they are more resistant to noise disturbance. For example, thermocouples help to monitor the temperature in the cutting zone [7], in the same way thermo – EDS vapors are measured, couples "tool - part ", allowing, according to the authors [8] "to get information about the pace of change in the dynamic behavior of the processing system due to tool wear." The method for measuring the electrical conductivity of the contact "tool - part" enters the group of integral methods [9]. Drawback of "integral" methods lies in their considerable inertia and the need to embed a thermocouple and electrical connections to the instrument.

Adaptive control presupposes monitoring tool status as well as surface quality control that is evaluated by the

results of the acoustic emission signals measurement and vibration.

The vibro-acoustic data signals turn out to be the simplest in the registration and subsequent processing [5]. However, in practice the use of these signals is related to a significant problem that is difficult to solve. These information signals are recorded by contact method. The implementation of the method requires a mechanical communication of the sensor with the source of the information signal (the surface of the object). To meet this requirement, it is necessary to solve two difficult tasks: to choose an informative point on the controlled objects where you want to install the sensor, and to avoid interference, always related to the method of measurement.

When monitoring the process of cutting an informative point, is considered, the point of the tool contact with the workpiece. At this point, the useful signal carrying information about the course of the cutting process arises. Placing the sensor at this point is impossible. Selection of other control points, the closest of which, for example, in turning the workpiece is on the cutting tool holder, makes it necessary to exclude the biggest interference from the measurement results

Noise disturbance in this case is the vibration generated by the machine constructive operating nodes. To select the useful signal without significant distortion on the background of the intensive level and complex in frequency parasitic vibration is almost impossible. The solution to this problem is achieved by eliminating the sensor contact with the vibrating surface of the machine tool that is implemented with non-contact measurement method.

The purpose of this article is grounding the rationale for the selection of a cutting sound as an information signal that is detected in a contactless manner.

III. Experimental Studies

A. Nature of the cutting sound

On the surface of the part resulting from machine tool processing irregularities are formed in the form of protrusions and depressions (Fig. 1), called roughness. The cause of the roughness is the trace left by a tool on the machined surface due to mutual oscillation of the tool and the workpiece. This phenomenon that is the tool interaction with the workpiece 135 years ago became the basis of recording, during which the cutter leaves a trail (a phonogram) on the lateral surface of the wax cylinder, and then on the surface of flat records.

Fig. 2 Represents a 1000 times magnified part of the sound track records, which is a part of the recorded track on a gramophone record. This soundtrack is just the surface of the part, which is formed while it is processed by a metalworking tool.

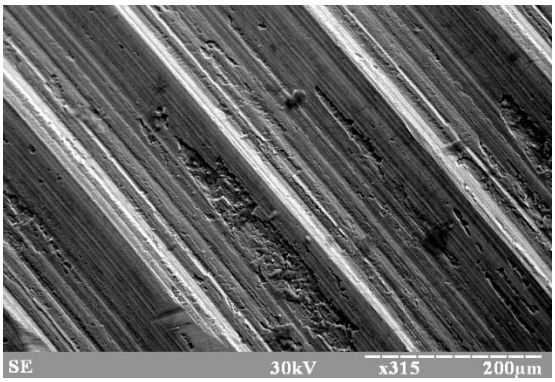


Figure 1. Increased by the microscope surface of the part after processing it by cutting

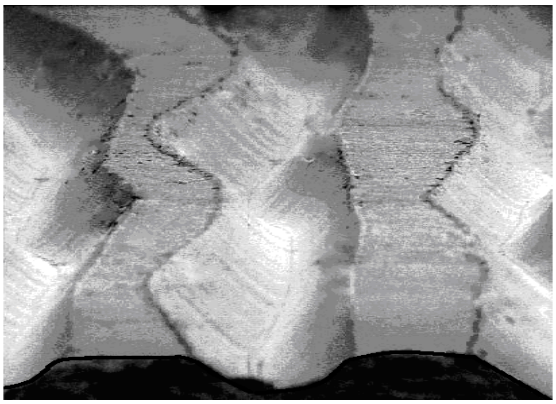


Figure 2. Increased by the microscope surface of the record

Roughness, similar to a phonogram contains information about the extent of tool wear and the quality of the processed surface of the workpiece, which is inextricably linked with tool wear.

A cutting sound is recorded by microphone typically of electret type (Fig. 3). The voltage of the electrical signal at the output of the microphone varies according to a change in the recorded sound.

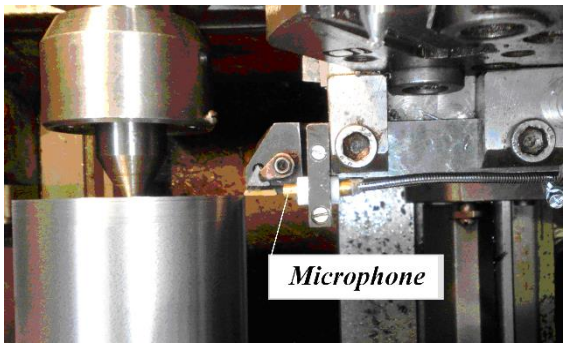


Figure 3. Installation example of microphone approximately the cutting area

Results of the studies revealed that the cutting sound has the properties of noise resistance, consequently it can

be considered as the initial information signal used in the adaptive control process of cutting. In particular, for this purpose the measured signal E_S is reduced to the dimensionless form $\left(\bar{E}_S = \frac{E_S(\tau)}{E_S(\tau_0)}\right)$, where $E_S(\tau)$, $E_S(\tau_0)$ – is an value of the sound, defined respectively in the current time and in the beginning of cutting process.

At the same time, the cutting sound should ensure solving the basic problem of adaptive control - maintaining the required in the part manufacturing documents quality, characterizing the degree of compliance of the part geometry and the purity of its surface with the drawing requirements. Geometric accuracy is determined by the tool wear and surface finish by the size of its roughness. The degree of correlation between the sound and these parameters has been studied experimentally.

B. The ratio between the tool wear and cutting sound

The reason for deviation of the part geometry from the details of the drawing is sized tool wear of the cutting tool h_r, equal in magnitude and opposite in sign of radius change R_∂ of the processed workpiece surface (fig. 4). The purpose of the experiments was to determine the correlation between the sound and tool wear while workpiece being machined on a lathe.

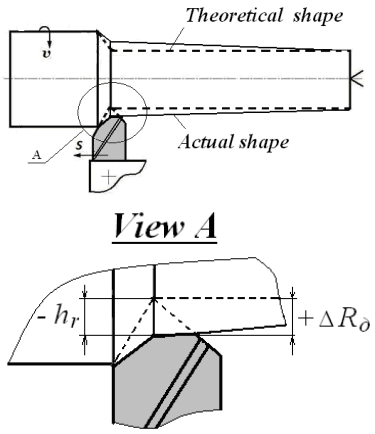
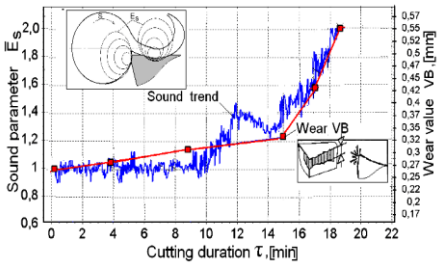


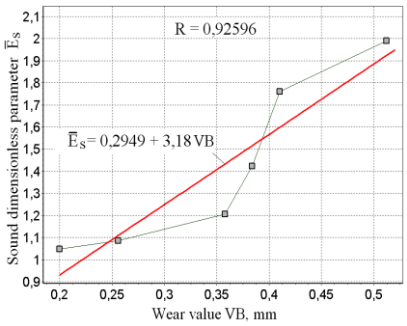
Figure 4. The deterioration of the geometric accuracy of the part because of the size wear of the tool blade

In the course of the experiments, the sound signal, was measured continuously. Measurements were carried out via a microphone installed near the cutting zone (Fig. 6) with the transmission of signal to the computer. Simultaneously, in steps (in two tool passes) using a measuring digital microscope, we recorded value of chamfer wear of major back surface *VB, мм* of the tool. The experiment was stopped, when the maximum allowable value of wear was reached.

The experimental results are shown in fig. 5, there is a graph of cutting sound change (parameter trend $\bar{E}_S$) and the curve of wear *VB*, as well as information about the correlation dependence between them, characterized by the correlation coefficient value *R*, equal to 0.926.



a)



б)

Figure 5. The relation between the curve of the tool wear and sound trend accompanying the treating by the cutting insert P25 of steel 12X18H10T for modes: *S* = 33 m/min, *f* = 0,15 mm/rev, *a* = 1,0 mm; a) the curve of wear *VB* and the trend of sound $\bar{E}_S$; b) the regression relationship between the dimensionless quantity of sound $\bar{E}_S$ and wear *VB*

As we see, the correlation of cutting sound and tool wear is big enough, that objectively indicates a high degree of coordination of change in the sound trend and wear curve.

C. The relationship between treated surface roughness and a cutting sound

The aim of the experiment was to determine the correlation dependence between the dimensionless sound parameter of $\bar{E}_S$ (1) and high-altitude R_a parameter characterizing the roughness. The roughness was measured at the longitudinal turning of steel billet St.40h on modes given in the caption to fig. 5. The roughness parameter *Ra* measurements were carried out periodically every five passes using a cutter-type profiler 283. The results of the measurements were recorded by the instrument dial indicator and additionally recorded on a laptop (fig. 6). The signal recorded on a laptop, subjected to further processing to determine the correlation between sound trends and roughness parameter.

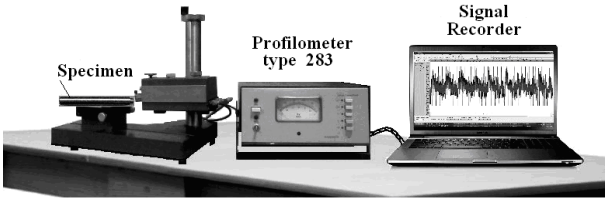


Figure 6. Registration roughness parameter *Ra* through profilometer type 283 and a laptop

The experimental results, are shown in fig. 7. In addition, for clarity and ease of comparison the roughness and sound roughness Ra , as well as the sound parameter $\bar{E}_S$, is given to the dimensionless form $\bar{R}a$ $\left(\bar{R}a = \frac{Ra(\tau)}{Ra(\tau_0)}\right)$, where $Ra(\tau_0), Ra(\tau)$ is an altitude roughness parameter defined respectively in the cutting process beginning and in the current time.Experiments have established that the correlation coefficient *R* between the sound trends and roughness parameter *Ra* equals to 0.944 (fig. 7).

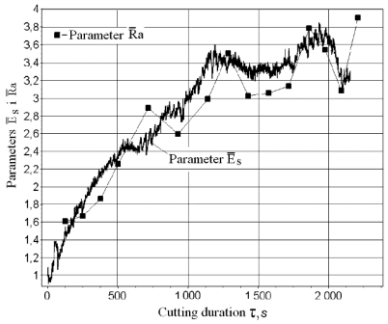


Figure 7. Comparison of sound trends (parameter $\bar{E}_S$) and roughness ($\bar{R}a$)

CONCLUSIONS

Thus, the results of studies presented in the article indicate the feasibility of using for the adaptive control of machining materials sound signal as initial information that is based on the following experimentally proven provisions:

– in contrast to the data signals, measured via contact method and subject, therefore, to noise disturbance generated by mechanical vibrations of the machine, a non-contact method of sound measurements via a microphone placed near the cutting area, and the transition in the processing of the measurement results to the dimensionless quantity of sound $\bar{E}_S$, provides noise resistant;

– sound trend and tool wear curve change during the cutting process with a high degree of coordination, as proven by a significant quantity of their mutual correlation coefficient *R* = 0,926;

– sound trends and altitude roughness parameter *Ra*, characterized by a correlation coefficient *R*, equal to 0.944.

Further research on the improvement of adaptive control of the cutting process should be directed at the development of quality forecasting methods of processing, allowing in the process of adaptive control to quickly determine the moment of the timely replacement of extremely worn-out tool. To solve this problem it is necessary throughout the entire workpiece treatment period monitor the cutting sound trend, quickly analyzing the nature of change and varies according to the analyses results cutting modes in order to extend the period of defect-free performance of a given technological operation.

ACKNOWLEDGMENT

The work was performed as part of the state budget scientific research theme of the Sumy State University, "Models and information design and management technologies in complex systems" (state registration number 0115U001569).

REFERENCES:

[1] Solomentsev Y. M. (et al.): The adaptive control of the technological process, 536 pp., Mechanical Engineering, Moscow (1980) (in Russian)

[2] Oraby S. E. Development of Models for Tool Wear Force Relationship in Metal Cutting / S. E. Oraby, D. R. Hayhurst // Int. J. Mach. Tools Manuf. – 1991. – № 33. – P. 125–138.

[3] Shao H. A cutting power model for tool wear monitoring in milling / H. Shao, H. L. Wang, X. M. Zhao // International Journal of Machine Tools & Manufacture, 2004. – February 2004. – № 44. – P. 105–108.

[4] Bolotin Yu. Z. Diagnosis life of the cutting tool / Yu. Z. Bolotin, V. Granovsky // Mashinostroitel. – 1989. – № 11. – P. 30. (in Russian)

[5] Litao Wang. Tool wear monitoring in reconfigurable machining systems through Wavelet analysis / Wang Litao, G. Mostafa Mehrabi Kannatey-Asibu Jr. Elijah // Engineering Research Center for Reconfigurable Machining Systems University of Michigan. – Ann Arbor, MI 48109–2125.

[6] Moriwaki T. A New Approach to Automatic Detection of Life of Coated Tool Based on Acoustic Emission Measurement / T. Moriwaki, M. Tobito // ASME J. Eng. Ind. – 1990. – № 112. – P. 212–218.

[7] Kretinin O.V. On the question of obtaining rapid assessment tool wear parameters of the fast / O.V. Kretinin, Klepin A.P., A.R. Quarters [et al.] // The theory of friction, lubrication and machinability of metals: works GPI.- Gorky, 1977. - Vol. № 4. - P. 112-121. (in Russian)

[8] Kretinin O.V. A possible approach to the assessment of contact phenomena at boundary lubrication / O.V. Kretinin, S.A. Kudryavtsev // Automated technology and mechatronic systems in mechanical engineering: Proc. scientific. works. - Ufa: USATU, 1997. – P. 152-159. (in Russian)

[9] Kourov GN Identification cutting temperature in the ACS process of turning on a lathe / G.N. Kourov, Yu. M. Kichko, I.I. Zubarev // Coll. scientific. works: Intelligent mechatronic machine systems.– Ufa, 2003. – 120 p. (in Russian)

Application for System Administrator of the Mobile Game

Ye. Pysmennyi, student; Yu. Parfenenko, seniour lecturer
Sumy State University
yuliya.parfenenko@gmail.com

***Abstract* – This article deals with the development of server side architecture of administrator of web-based game application. A web-based application with implemented features for management of server side of administrator was built.**

***Key words* – web-based application, game, system administrator, client, server, data flow, control.**

I.INTRODUCTION

Web-based game applications consist of a client side and server side. The client side is responsible for a user interface and game server connectivity, and the server side describes the application logic and represents powerful features of a quick and secure user interaction with each other, the involvement of local and "cloud" medium for storage of required data etc.

Such web-based game applications are characterized by hierarchical user role distribution, different levels and tools of run through the game, the physics of moving and placing of objects with textures, units of remuneration, mission, language etc.

There is a need to administer a data flow of the game application. That is, to be able to control dynamic elements of the game and control the correctness of the server features, without the intervention of a web-application code or directly in a database.

The project aim is to develop the server side of administrator of web-based game application with the ability to control the data flow between client and server on the website.

The server side of administrator will be able to test the existing or new features of the game, to implement CRUD – control of roles, news and languages of the game as well as to control XML and JSON data formats. This project has a functional value, since its capabilities provide the performance of tasks to control the game by administrator.

II. A ANALYSIS OF DEVELOPMENT PROBLEM OF ADMINISTATIVE SIDE OF WEB-BASED GAME APPLICATIONS

Today, the majority of high-loaded web-based game applications use special data flow control systems of own servers for administrators. But such server sides serve only for a few responsible people who administer the game. That's why they are locked for use by all other users. Such systems are developed for specific tasks and there are no clear common rules that must be met by created features.

Distribution of operation of high-loaded web-based game applications are an integral part when planning and implementing the architecture of their operation. As this criterion has influence on:

- high-speed-response;
- security of user interaction;
- usage the local and "cloud" medium for storage of data;
- performance;
- usability;
- user interface etc.

Once the web-based game application is downloaded by a client (or is opened in a browser), the client begins to interact with its interface. Thus, any dynamic changes are sent to the server on which the received data are processed and after this the processed data are sent back to the client. In turn, the user gets the result in their device or browser.

To encourage the users to use the web-based game applications, scripts and storylines are created, various types of user roles, levels and tools of run through the game, motion trajectory and location of objects with textures, units of remuneration, mission are developed. All these attributes are dynamic attributes for each user. Their logic is described on the server and data about them are stored in a database. Therefore, the possibility to administer these data flows is very important and relevant task.

A similar problem occurs in the IOS strategy game "Battle of Ages". It is a mobile game application. A type of gameplay is similar to a real-time strategy with the expectation of a large number of users.

The game has a server that provides its logic written in PHP, the main database located on Amazon RDS and server for IOS - developers. The gameplay is based on the control of their own game objects for the development of a city, collection of achievements, passing missions, conducting of wars.

The main purpose of a player is to run through the game scene by forming a military and economic power and making strategic decisions on the development of the plot.

For this game the new features are gradually added. These features should be checked for correctness of operation and should able to operate with data formats such as XML and JSON [1].

It is necessary to develop the mobile game administrator application which results in absence of necessity to interfere in web-based application code [2].

This admin panel is a special management system of data flow between the client and the server.

This will allow making the process of manipulation of the game faster and more efficient by CRUD-control and testing the existing or new features of the game, if necessary.

III. THE SELECTION OF METHODS FOR DEVELOPMENT OF MOBILE GAME ADMINISTRATOR APPLICATION

For development of server side of administrator of web-based game application a programming language and database are used; this database can quickly handle requests and maintain high load. Of the available options the most optimal solution is PHP and database MySQL.

The main methods of implementation are structured programming and object-oriented programming [3].

When creating the medium sized applications the structured programming is used, the idea of which is that the structure of the program should reflect the structure of the original problem in order to make a solution algorithm clearly evident from the text.

This approach is convenient because it allows a person to think on the object level, not use specific operators and variables. However, this method has a drawback - in structured programming any modification leads to changes in the logic of the program. Thus, the development of server side of web-based game application administrator can be not user friendly and can take a lot of time.

Object-oriented programming (OOP) is the method of program creation, which is based on the concept of the class as a structure, which describes a set of similar real-world objects and their behavior. The task which is solved by the technique of OOP is described by terms of classes and operations performed on objects of that class.

The program in this approach is a set of instantiated objects and the relationships between them. The data compared with the processes are more stable and relatively rare are changed by the system part. Hence the main advantage of the object-oriented approach is: object-oriented systems are easier to change, because their construction is based on the stable forms. This enables the system to develop gradually and does not lead to its complete rebuilt even if there are significant changes in output requirements.

OOP is widely used for frameworks that facilitate the development of web-based applications.

The framework YII2 is used for this project. Its main advantages are:

- integrated generator queries to database;
- ActiveRecord for relational and NoSQL database;
- RESTful API;
- multilevel caching support;
- easily extensible for new features;
- powerful RBAC system;
- uses standard methods of problem solving that reduces or eliminates the complexity of the code;
- integrated validation;
- facilitates teamwork by means of compliance with standards (PSR-4);
- facilitates support code using general architecture and methods;
- GII generation module of models, controllers and modules.

It is based on MVC pattern [4].

IV. ARCHITECTURE OF ADMINISTRATOR APPLICATION OF MOBILE GAME

Server side of administrator is implemented using MVC pattern. This pattern is appropriate for creation complex projects and has become a popular strategy for building web-based application. MVC separates user interface, data and processing of user actions into three components: Model, View and Controller [5].

Model handles the state of the application.View is the representation of the user interface. Controller receives user requests, and translates them into actions that the Model should take. Then it selects the appropriate View to handle the response [6].

MVC pattern represents the possibility to be integrated with additional configurations, modules etc. [7]. This separation gives an advantage that the changes in a separate component of the system will not affect the entire logic of the server side of the game.

Game server already uses MySQL. MySQL database contains information on players, cities, weapons, game levels, number of awards and tools. Also, it stores the data of role hierarchy to enter the system being developed [8].

Component View is responsible for displaying of client side of administrator. That is, it describes a user interface that is accessible through a web-browser on website.

Controller accepts requests (depending on the actions of the administrator) on the server. This component is responsible for the logic distribution between the data processing in the Model and display in the View.

Model receives data from the Controller. It is in this component the features of CRUD-control (Create, Read, Delete, Update) of news, roles and languages of games, the control of data in XML and JSON formats, distribution of user roles is implemented. Model has a direct connection with MySQL database and connects GII module and RBAC system for role control.

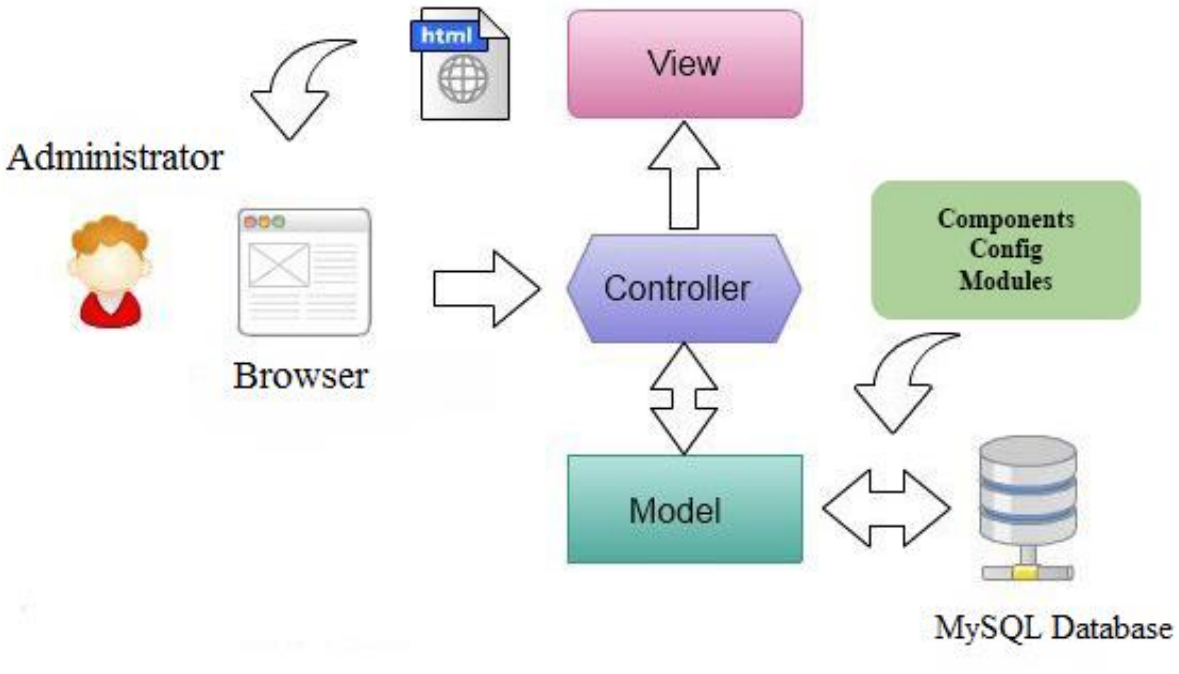


Figure 1 – System Architecture

To enter the system, you have to authorize, after which the features of server side of administrator will be available.

The administrator will be able to do the following:

- load and parse the file of XML format on the server
- convert the files from XML to JSON
- load JSON data into database tables
- appoint a new administrator of the game
- perform CRUD - control of news and languages of the games
- conduct a test checking of the existing or new features of the game if necessary
- set an access control for individual users of system being developed.

The role distribution is provided by RBAC - system (Role - based access control), which is installed as a separate extension into the server side [9].

System implements a general hierarchical RBAC interface for user, following the NIST RBAC model. It provides the RBAC functionality through the AuthManager application component. A role represents a collection of permissions. To check if a user has a specified permission, we may check if the user is assigned with a role that contains that permission.

Using RBAC involves two parts of work. The first part is to build up the RBAC authorization data, and the second part is to use the authorization data to perform access check in places where it is needed.

The administrator has the ability to control access to the system, assign new users and give them individual permissions of server side of administrator panel.

CONCLUSION

The created architecture of server side of administrator of web-based game application will save money and time by flexible and atomic work of interconnected features. This system enables the dynamic control of elements of the game and control the correctness of features of the server without the intervention of a web-application code

REFERENCES

[1] Poleznyie servisyi dlya veb-razrabotchika [Electronic resource]- Access mode: http://zinchenko.us/article/ru-useful-services-for-developers/
[2] Veyl, Estel HTML5. Razrabotka prilozheniy dlya mobilnyih ustroystv / Estel Veyl. - M.: Piter, 2014. – S. 661.
[3] MySQL php connect [Electronic resource]- Access mode: http://webcheatsheet.com/php/connect_mysql_database.php
[4] YII2 project basic template [Electronic resource]- Access mode: https://github.com/yiisoft/yii2-app-basic
[5] Patternyi MVC [Electronic resource] – Access mode: https://habrahabr.ru/post/215605
[6] The Model View Controller pattern in web applications [Electronic resource] – Access mode: https://zeekat.nl/articles/mvc-for-the-web.html
[7] Realizatsiya MVC [Electronic resource] - Access mode: https://www.youtube.com/watch?v=ba3M3_Myrqg&list=PLSdH7dYnlGYgQ6ElbHRpG2zRXbkgkO3zQ
[8] Velling, L. Razrabotka veb-prilozheniy s pomoschyu PHP i MySQL / L. Velling. - M.: Dialektika / Vilyams, 2016. – S. 794
[9] Web role control [Electronic resource]- Access mode: http://www.yiiframework.com/doc-2.0/guide-security-authorization.html#rbac

Development and Research of Probabilistic Models of Quality Assessment of Management Information Systems Operation

A. Litvinov
O.M. Beketov National Universityof Urban Economy in Kharkov, litan@meta.ua

Abstract - The purpose of the research is the development of stochastic models of management information systems (MIS) operation based on queueing systems. It has been shown that it is possible to use a single line queueing systems with the generalized erlang flow of random events. The study of such systems is carried out and the main characteristics of operation are obtained. This allows the planning of procedures for MIS operation.

***Key words* - time, criterion, *model,* service, process, *design, flow,* system, event.**

I.INTRODUCTION

MIS are difficult to implement, their work is often associated with numerous and mutually independent streams of input events, wearing casual character. In this regard, on the design phase of such systems the use of stochastic modeling is necessary [1, 2].

II. THE MAIN PART

Software of MIS provides the control of wide range MIS individual sub-system tasks solution in real time. These tasks are characterized by frequency of execution, duration of work, priority, volumes of output and input information. The launch on the execution of various tasks carried out by MIS supervisor system on the commands, which the planner of MIS periodically prepares [3]. The simple form of planning of the real time tasks execution is synchronous, when slot between sequential starts of the same task are equal to the selected slot. The flexibility and efficiency of such planning is greatly improved with the introduction of priorities, as well as with the ability to solve the background tasks when the computer system of MIS is free from solving real-time tasks [4, 5]. A variety of characteristics can serve as a criterion of the effectiveness of the dispatching operation of the system. For example the delay of start time of individual tasks on the planned time: the mean time, the probability that the time delay exceeds a certain fixed value and so on [6]. The knowledge of these characteristics can be used for synthesis of the system of planning of MIS software.

Let the whole time interval is split with the timer into separate intervals Δ. For each interval a number of high-priority and low priority tasks in accordance with the operation of MIS is planned to perform. Time left free in the interval from solving real-time tasks is used to solve background tasks (for example, test equipment monitoring).Interruption of low priority tasks is authorized.

Interrupted low priority task complete service after processing of high priority tasks. If a low priority task is interrupted in *i* - th interval, it continuation is added to the execution of tasks in (*i* + 1) - th time interval. Each task perform is planned with a certain period T_i. You can select the main computing cycle ***T***, which is the least common multiple of T_i, $i = 1,2,...,N$, where *N* - total number of individual tasks, that can be solved with the help of MIS software. In general, the time of execution of each task can be random; therefore it is important to know their solving time delay. This is especially true for low priority tasks, the beginning of the implementation of which can be shift by high priority tasks.

For considering execution of high priority tasks assume the next. Let ξ_{ij} is the time of execution of *i*-th priority task in the *j*-th interval of time. Suppose that $\Phi_j = \sum_{i \in I_j} \xi_{ij}, i = 1,2,...,n,n+1,n+2,$ is the time of execution of all high priority tasks in the *j*-th interval of time, where I_j are scheduled tasks on the *j*-th interval, and $n = T_h / \Delta$, where T_h is the main cycle of execution of high priority tasks. Let φ_j is the average time of execution of priority tasks in the *j*-th interval. The overall distribution function of service time can be written as a superposition of flows for each service interval - $B(t) = 1 - \frac{1}{m}\sum_{j=1}^{m} e^{-\mu_j t}$, so it is described as hypereksponentsial distribution. Input stream for such a system is regular, but it can be approximated by Erlang distribution in the event that the coefficient of variation approaches to zero. So, the entire system can be represented as a single line queueing system of type E/H/1. Let us investigate this system. Let

$\bar{\lambda} = \frac{\lambda}{k}$ and $\bar{\mu} = \left[\sum_{i=1}^{m} \frac{1}{m\mu_i}\right]^{-1}$ are average intensity of the flows described in accordance with the distribution of $A(t)$ and $B(t)$. $\alpha(s)$ and $\beta(s)$ are the Laplace-Stieltjes transformation of these distributions. So

$$\alpha(s) = \left(\frac{\lambda}{\lambda + s}\right)^k \text{ and } \beta(s) = \frac{1}{m}\sum_{i=1}^{m} \frac{\mu_i}{\mu_i + s}.$$

If $\bar{\mu} > \bar{\lambda}$, then for a stationary mode exists a function of the distribution of the waiting time $F(t)$, for which it is possible to compile the Lindley integral equation [7]. It corresponds to the factorization equationon

$$\gamma(s) = \frac{K_+(s)}{K_-(s)} = \alpha(-s)\beta(s) - 1 = \\ = \frac{1}{m}\left(\frac{\lambda}{\lambda - s}\right)^k \sum_{i=1}^{m} \frac{\mu_i}{\mu_i + s} - 1 = 0 \quad (1)$$

Notice, that function

$$\beta(s) = \frac{\frac{1}{m}\sum_{i=1}^{m} \mu_i \frac{\prod_{j=1}^{m}(\mu_j + s)}{\mu_i + s}}{\prod_{i=1}^{m}(\mu_i + s)} = \frac{P_{m-1}(s)}{Q_m(s)} \quad (2)$$

is the ratio of two polynomials, and the degree of the numerator is less than the denominator. Then the Laplace-Stieltjes transformation from $F(t)$ will be

$$\varphi(s) = \frac{Q_m(s)}{Q_m(0)\prod_{i=1}^{m}\left(1 - \frac{s}{q_i}\right)}, \quad (3)$$

where $q_i\,(i = 1,2,...,m)$ are the roots of equation

$$\gamma(s) = 0 \Rightarrow \frac{1}{m}\left(\frac{\lambda}{\lambda - s}\right)^k \sum_{i=1}^{m} \frac{\mu_i}{\mu_i + s} - 1 = 0, \quad (4)$$

which are located in the left half-plane $\operatorname{Re} s < 0$. Using (2), we find in the left half-plane $\operatorname{Re} s < 0$

$$\varphi(s) = \frac{Q_m(s)}{\prod_{i=1}^{m}\mu_i\left(1 - \frac{s}{q_i}\right)}. \quad (5)$$

Let $s = \lambda(1 - z)$, then the equation (4) will have the form

$$\frac{1}{m}\sum_{i=1}^{m} \frac{\mu_i}{\mu_i + \lambda(1 - z)} - z^k = 0. \quad (6)$$

Solving this equation after the transformations, we will get

$$\varphi(s) = \lambda^m \prod_{i=1}^{m} \frac{(\mu_i + s)(1 - z_i)}{\mu_i + [\lambda(1 - z_i) - s]}, \quad (7)$$

where z_i are the roots of equation (5) and

$$1 < z_1 < 1 + \frac{\mu_1}{\lambda} < z_2 < ... < 1 + \frac{\mu_{k-1}}{\lambda} < z_k < \\ < 1 + \frac{\mu_k}{\lambda} < ..... < z_m < 1 + \frac{\mu_m}{\lambda}. \quad (8)$$

Taking the inverse Laplace-Stieltjes transformation from (7), we can obtain an expression for the distribution function of delay time in serving high-priority tasks $F(t)$.

Let's find the average delay time in the service of high-priority tasks

$$\bar{t}_d = -\varphi'(s)\big|_{s=0} = \sum_{i=1}^{m}\left[\frac{1}{\lambda(z_i - 1)} - \frac{1}{\mu_i}\right]. \quad (9)$$

)

Probability that time of solving of high-priority tasks ξ will exceed Δ is determined by the next expression

$$P\{\xi > \Delta\} = \sum_{j=1}^{m} e^{-\mu_j \Delta} \Big/ m \quad . \quad (10)$$

Real-time conditions have a strong impact on the execution of lower priority tasks, for which the interruption of service is allowed [6]. Suppose that *i* is the priority level and $T_i = N_i / \Delta$ the period of performance of the low priority task, which are scrutinized. Let u_j is the time required to execute this task in the *j*-th interval, and ω_j is the total time of execution of all high priority task in this interval including the time for the end of execution all tasks from (*j* - 1)- th interval. Then the time that remains in the j-th period for the execution of low priority task is defined by the expression $v_j^i = \max\{0, T_i - w_j\}$, where v_j^i is a random variable. It can be either greater or less than the time u_j required to solve low priority task. Thus, the relationship between these random variables is similar to the process of service of requests in a single-server queueing system of type G/G/1; the sequence of values v_j^i describes the

input stream of requests, and u_j - service time. To determine the relationship between the delay time Z when executing a low-priority task and the waiting time in an equivalent single-line queueing system, we will use the fact that if the solution of the problem starts in the planned time interval Δ, then $\eta = z$, otherwise $\eta = z - \Delta \cdot \text{int}(z/\Delta)$, where $\text{int}(x)$ is the integer part of the expression. So

$$P\{z > t\} = P^*\{\eta > t^*\} = 1 - F(t - \Delta \cdot \text{int}(t/\Delta)) =$$
$$= 1 - W(t - \Delta \cdot \text{int}(t/\Delta)),$$

Where $F(t)$ - is the law of the distribution of the waiting time for an equivalent single-line queueing system, and $W(t)$ a is the distribution function of the waiting time.

Let us approximate the input stream and the service flow with distributions that belong to the family of common Erlang distributions. The Laplace-Stieltjes transformation from their distribution functions have the form

$$f(s) = \prod_{i=1}^{k} \sum_{j=1}^{n_i} \frac{a_{ij}\lambda_{ij}}{s + \lambda_{ij}}, \quad \sum_{j=1}^{n_i} a_{ij} = 1, \qquad (11)$$

and is the ratio of two polynomials.

You can get comprehensive performance characteristics of low priority tasks in real time if you will approximates the input flow and service time by general Erlang distribution. Exponential, hyper-exponential and Erlang distribution are special cases of distribution (11). Any distribution of the duration of intervals between neighboring events can be approximated with any accuracy by the general Erlang distribution, it is only necessary to choose the parameters appropriately $a_{ij}, \lambda_{IJ}, n_i, k$. It is desirable that the values n_i and k will be small.

Let us consider a single-line queueing system with expectation that simulates the process of servicing low priority tasks, the input flow of requests and services for which are given by generalized Erlang distributions with Laplace-Stieltjes transforms from the distribution functions, respectively $\alpha(s)$ and $\beta(s)$.

The expression for the average intensity of requests for low-priority tasks is calculated as follows:

$$\overline{\lambda} = -\alpha'(0)^{-1} = \left[\sum_{i=1}^{k} \sum_{g=1}^{l_i} \frac{a_{ig}}{\lambda_{ig}} \right]^{-1}. \qquad (12)$$

The expression for the average intensity of the process of servicing low-priority tasks will have the form

$$\overline{\mu} = -\beta'(0)^{-1} = \left[\sum_{i=1}^{n} \sum_{q=1}^{c_i} \frac{b_{iq}}{\mu_{iq}} \right]^{-1}. \qquad (13)$$

If $\overline{\mu} > \overline{\lambda}$, then for the system, there is a steady-state regime, for the study of which the above approach can be used

III. CONCLUSIONS

MIS operate under random event flows. In this regard, on the design phase of such systems the use of stochastic modeling is necessary, in particular on the basis of queueing systems. In paper process of planning of real time application is analyzed. The planning cycle is composed of solutions of high-priority and low-priority tasks. It has been shown that it is possible to use a single line queueing systems with the Erlang input stream and hyperexponential service time for the simulation of high priority tasks. The integral equation Lindley for such a system is received. The solution of such equation is obtained by factorization of integral equation. Explicit expressions of the distribution function of the delay time of the real-time high-priority tasks is received, as well as its average value. The process of execution of low-priority tasks is analyzes. It is shown that the analysis of this process can be done by the use of queueing systems with the generalized Erlang flow of random events. On the basis of the method of the integral equation Lindley the Laplace-Stieltjes transform of the waiting time in the queue and the average value of this time is obtained.

REFERENCES

[1] V.I. Boltenkov, A.L. Litvinov, i N.V. Lycheva, N.V. Konfigurirovanie i nastrojka avtomatizirovannyh informacionnyh sistem. Belgorod: BelGU, 2005.

[2] V.M. Trojanovskij. Informacionno-upravljajushhie sistemy i prikladnaja teorija sluchajnyh processov. Moskva: Gelios ARV, 2004.

[3] V.B. Kropivnitska, B.L.Klim, A.G. RomayuIuk, M.O. SladInoga.”DoslIdzhennya algoritmIv dispetcherizatsIYi v komp'yuternih sistemah”, RozvIdka ta rozrobka naftovih I gazovih rodovisch, no. 2(39), c. 93 – 105, 2011.

[4] . Goma. UML Proektirovanie sistem real'nogo vremeni, parallel'nyh i raspredelennyh prilozhenij. Moskva: DMK-Press, 2011 .

[5] Rajib Mall. Real-Time Systems: Theory and Practice. IGI Global, 2006.

[6] A.M. Sulejmanova. Sistemy real'nogo vremeni. Ufa: UGATU, 2004.

[7] G.P. Klimov. Teorija massovogo obsluzhivanija. Moskva: MGU, 2011.

Information Technology for 3D Relief Model Generation By Color Parameters of Pixel

Yan Voitsekhovskyi, Iryna Baranova
Sumy State University, baranova.iryna@gmail.com

***Abstract* – Using of digital maps in simulators will dynamically change the tasks options and improve the quality of military professionals. The paper said information technology to get 3D relief model for digital maps.**

Created simulator module has a user friendly interface with a large number of configuration options that allow well prepared for raster images for recognition. The generation of digital terrain automatically. The resulting digital map can edit, change display settings, zoom and move in the world coordinate system, add it necessary to vegetation and buildings.

***Keywords* – information technology, bitmap image, color pixel, algorithm, method, height map, 3D relief model.**

I. Introduction

The defensive capacity problems in Ukraine require a high-quality training of military professionals. In this regard, there is a need to improve the professional military education.

Implementation of modern information technologies into the military specialists training is unquestioned. One of these process components is the use of various interactive simulators in the training process, so their development is a topical task.

Thus, it is possible to formulate the main purpose of work - the creation of information technology for 3D relief model generation by color parameters of raster image pixel

The developed technology allows to carry out the interactive generation in the real time topographic maps in the form of 3D relief models for use in the military preparation simulators.

To achieve the goal, it is necessary to solve the following problems:

- to analyze existing methods and analogues;
- to formulate functional requirements for the module simulator;
- to develop technology of generating digital terrain;
- to develop a relief structure generating module;
- to implement a program part of the module.

II. The Analyze Existing Methods and Analogues

An analysis of the existing methods for generating a digital terrain model has been considered in detail in this paper [1].

As to available data, popular and widely distributed generation software tools are almost absent among domestic developments. There are several software's that are designed to solve highly specialized issues but do not have a full value generation module and cannot be used to solve this problem:

- Populate: Terrain in 3ds Max [2];
- TERETORY – created in SSU, military training department.
- Three-dimensional map representation in Digitals software [3];
- Relief generating using Unity's Terrain system [4];

The functional requirements to which the created module simulator should correspond were formulated:

- loading and editing (rotation, movement, zoom) the most common raster image formats;
- formation of the digital relief map (DRM) and its texturing;
- automatic creation of additional objects (grass, trees, rocks, buildings);
- the DRM format should be convenient for further use in the simulator development software;
- export of the digital relief map.

Military department of Sumy state university has developed the TERETORY simulator. For achieving the set task, the existing structure were added 3 points that improve the simulator:

- Generating of a 3D model in the color image elements.
- Transfer of the generated territory 3D model;
- Loading of the received territory 3D model;

It considerably improves the process of generating 3D - surfaces.

III. Structural and Functional Analysis Technology

By using SADT methodology was conducted structural and functional analysis technology, presented on the Fig. 1 and Fig. 2 in the form of contextual diagrams and charts decomposition.

- loading
- editing;
- formation of the digital relief map;
- texturing;
- automatic creation of additional objects;
- relief map format should be convenient for further use in the simulator development software;
- export of the digital relief map.

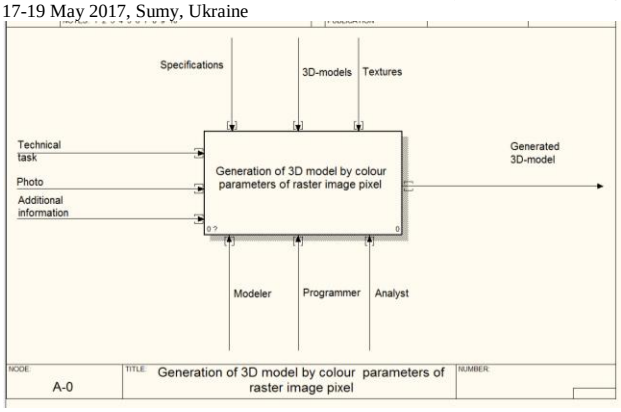


Figure 1. Contextual diagrams of technology

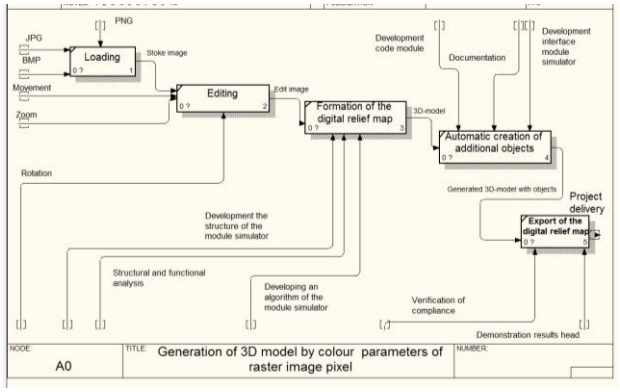


Figure 2. Diagrams of charts decomposition

IV. TECHNOLOGY OF COLOR MAPPING

For practical implementation of the task as the base was used plugin game engine Unity [5], which was integrated module designed simulator. The connection between the base and plug-in module is shown on the Fig. 3.

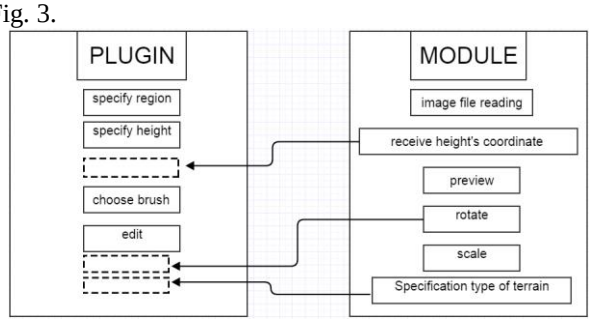


Figure 3. Communication between Unity 3D and the module

Usually in forming the relief is used [6]:

- the first – to use of a regular height grid (or its another name - HeightMap).
- the second – to use of an irregular height grid and communications that connect them (i.e. storage of a simple triangulated map).
- the third - a landscape map storage, but in this case no specific height is stored, but the information about used blocks. In this case a certain amount of pre-built segments is created, and only indexes of these segments are indicated on the map.

In this research only landscapes with a regular grid will be considered. It will allow to use rather simple algorithms for their generation. When using a regular grid (elevation map) data presented as a two-dimensional array (Fig. 4). Specified two coordinates (x, y - height and width array), and the third coordinate given value in a particular cell is the height.

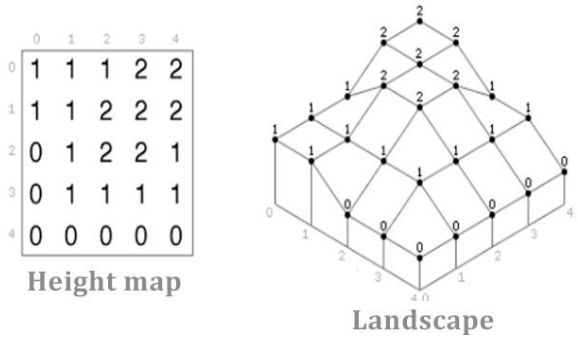


Figure 4. Converting a two-dimensional array in relief

Typically elevation map stored in image files (Fig. 5). This makes it easy to change and more or less clearly view the data. Then two coordinates to position a particular pixel in the image, and the third coordinate will be represented by color (direct dependency on brightness color components of the pixel, the higher the brightness - especially height values for this point). Typically, these images are in black and white version, but you can use all the colors. The second option gives us more height gradations than 256 gradations predictable if monochrome presentation.

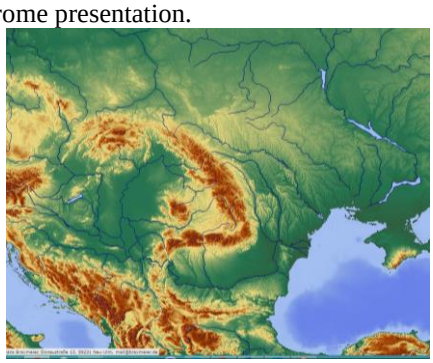


Figure 5. Image file of map

To form the two elevation map coordinates x, y choose the size of bitmap (pixel current position horizontal and vertical of image). The third coordinate (height) will be determined by the color map, which displays color scalar data. Scalar values serve as indices in the reference color (Fig. 6) lookup table [7].

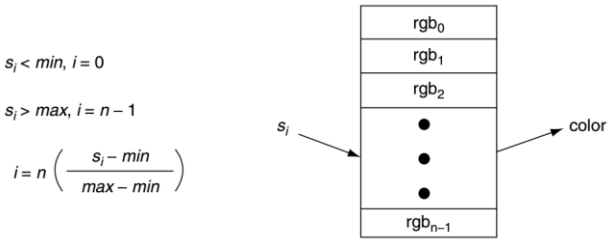


Figure 6. Displaying a scalar colors through a table

The table contains an array of colors (red, green, blue (RGB) components or other similar views). The indexes are defined as follows. Scalar values greater than the maximum value range, color shifts to the maximum scalar value and lower minimum range - shifting to a minimum of color value.

The resulting data is transmitted in Unity for future generation of relief using embedded objects Lines, Polygons, Mesh and displayed on the screen.

V. PRACTICAL RELEASE

Before working in Unity, a new project is created and a module area is marked. At the same time the standard built-in tools of the program are used.

Module import to the project happens with Drag & Drop method. Then the developed module is loaded by means of the appropriate command in the Window menu or a specially tuned combination of the "hot" keys Ctrl + G.Work module begins with loading the saved file raster maps. Further verification comes loaded maps correctness, defining attributes of colors and sizes maps. Then apply existing texture image generation performed main texture filter checks for and receiving component filter surface. Next generation topography performed with a given dimension (width, height, depth, resolution) and the result is a surface.The simulator created module has the convenient interface with configuration setups that allow to prepare qualitatively raster images for further recognition. Generation of a digital relief is carried out automatically. Terrain map can be displayed in color as form and in tinted form and a polygonal mesh, which allows you to check the correctness of topology construction (Fig. 7).The received vector map can be edited; display parameters, scaling and relocation in a world coordinate system can be changed, with adding to it all the necessary objects of flora and buildings.

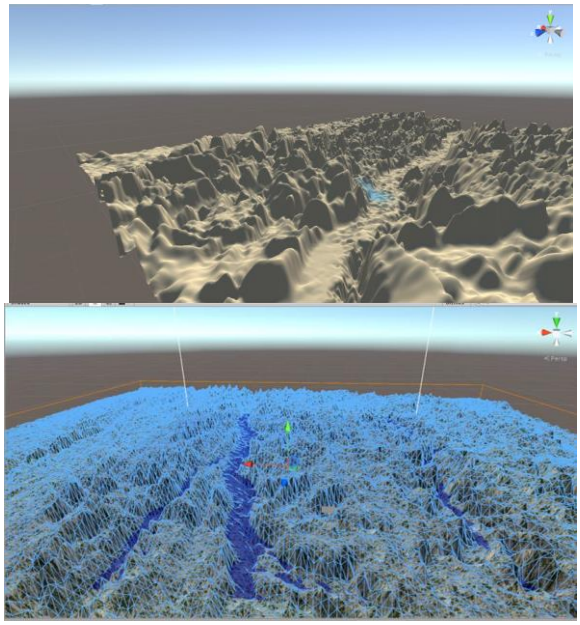


Figure 7. Options terrain map display

VI. CONCLUSIONS

As a result of work these problems have been resolved:

- analyzed existing algorithms for obtaining digital maps and software for generation of terrain;
- formulated functional requirements generated by the module simulator;
- developed the technology of the module simulator based on selected methods and tools for generation of terrain;
- developed structure and algorithm module simulator that allows dynamically generate digital map real relief;
- realized module simulator among the game engine for future use when creating a simulator as a whole.

Developed module allows you to generate real options challenges and improve quality of training the military.

REFERENCES:

[1] Y.Voitsekhovskyi, V.Zubko, I.Baranova. The technology and construction methods of digital terrain models/ IVth International Conference «Advanced Information Systems and Technologies, AIST 2016» – Sumy, 2016 – 130 pp. (in Ukrainian)

[2] Populate:Terrain [E-resource] –: http://populate3d.com/ products/ terrain/.

[3] Software for creating digital maps «Digitals» [E-resource] – http://www.vingeo.com/Rus/podigitals.html. (in Russian)

[4] Terrain Engine [E-resource] – https://docs.unity3d.com/Manual/script-Terrain.html.

[5] Introducing GAIA. The Terrain Creation Tool For Unity 3D [E-resource] – http://www.procedural-worlds.com/gaia/.

[6] Generation of three-dimensional landscapes [E-resource] – http://www.ixbt.com/video/3dterrains-generation.shtml. (in Russian)

[7] Schroeder, Will; Martin, Ken & Lorensen, Bill. *The Visualization Toolkit* (4th ed.). Kitware, 2006.

SESSION 5
INTELLIGENT DATA PROCESSIN

The task of schedule optimizing for partially ordered jobs on machines with different productivity in the presence of idle time

Raymond Keller, Olesia Ilchuk
National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", raybed@ukr.net

***Abstract* – The subject of this article is the job shop scheduling problem and methods for solving this problem. It contains model for the task of schedule optimizing for partially ordered jobs on machines with different productivity in the presence of idle time and algorithm for getting initial permissible solution.**

***Keywords* – job shop scheduling; optimizing; initial acceptable solution; algorithm; theory of schedule.**

I. INTRODUCTION

Scheduling theory deals with "ordering" and "scheduling". Ordering - it queuing operations performed by a machine, and scheduling - a task action sequences for several cars. In this scientific paper the "scheduling", namely the class of problems in which we have a service system that consists of machines. Each machine performs (some basic tasks). These operations are partially ordered, that perform one operation may depend on a previous execution did not depend. However, the duration of the operation is not a constant, it depends on the machine on which it runs. At the time axis may be "gaps" that forced the machine idle periods, when it cannot perform any operations. The set of machines, transactions and operations disciplines relevant purpose machines called maintenance. Scheduling this process means that for every transaction on the time axis corresponds to the subset in which the operation is performed in accordance machine. The main objective of this class of problems is to reduce maintenance time.

II. EXISTING APPROACHES TO SOLVING PROBLEMS SCHEDULING

A. Local Search

Local Search - search carried out by local search algorithms, a group of algorithms which search is conducted only on the basis of the current state and earlier passed states are not counted and not memorable.

The main purpose of the search is not an optimal path to the target point, and some optimization objective function, so the problem solved similar algorithms, called optimization problems. To describe the state space such problems using state space landscape, this presentation of the problem is to find the state of global maximum (or minimum) at this landscape.

B. Branch and bound method

Algorithmic method for finding the best solutions of various optimization problems, in particular, NP-complete problems, such as the problem of traveling salesman. Method is a combinatorial (sorting algorithm) with the rejection of subsets of feasible solutions that do not contain optimal solutions. The idea of the method can be described by the example of finding a minimum or maximum of a function f (x) on the set of admissible values of x. The method comprises two procedures: branching and evaluation (search boundaries).The first is to partition the set of feasible solutions to smaller subsets (technology "divide and conquer"). The procedure can be repeated recursively for all subsets. The result will be obtained tree is called a tree branch and bound tree or search nodes which are received subset. The assessment procedure is to find upper and lower bounds for the optimal value on a subset of feasible solutions.Screenings decision is made in the following way. To minimize problems, if the lower limit for the subset A search tree than the upper limit of any of the previously discussed subsets of B, then A can be eliminated from further consideration. The smallest of the received top ratings memorize in the variable m. Any node tree search, the lower limit is greater than the value m, can be excluded from further consideration. If the lower boundary coincides with the node tree top, this value is the minimum of the function at the appropriate subset.

C. Genetic algorithms

Genetic Algorithms - adaptive search, which recently often used for solving functional optimization. They are based on the genetic processes of biological organisms, biological populations develop over several generations, obeying the laws of natural selection and the principle of "survival of the fittest".

The task is coded so that its solution could be represented in the form similar to the array data structure of the chromosome. This array often called just that: "chromosome". Random array creates a certain amount of initial elements "Persons" or initial population. Individuals are valued using the tool accessories, which resulted in each person is assigned a certain value fitness, which determines the possibility of the survival of the individual. Then using obtained values adaptability selected persons admitted to crossing (selection). Persons

used "genetic operators" (in most cases an operator intersection (Crossover) and mutation operator (mutation)), thus creating the next generation of people. Those next generations also evaluated using genetic operators and performed selection and mutation. Yes, simulated evolutionary process continues several life cycles (generations) until stopping criterion satisfied algorithm

Genetic algorithm consists of several stages:

1. creation of initial population (chromosome create for each individual)
2. the calculation of each individual fitness function (that it shows who fit best in this population)
3. Selection of the best representatives for further education of offspring
4. crossover
5. mutation
6. When (4) get the children, some of which passes through (5). The output obtain offspring
7. Selection of parents and children in the next generation
8. return to step (2) if the values that give our children is not satisfied

D. Simulated annealing

Simulated annealing - general algorithmic method of solving the problem of global optimization, especially in discrete and combinatorial optimization, in which the procedure for finding global solution simulates the physical process of annealing.

Exotic names associated with the algorithm of simulation methods in statistical physics, based on the technique of Monte Carlo. Investigation of the crystal lattice and behavior of atoms slow cooling of the body led to the appearance of the light probabilistic algorithms that have proven extremely effective in combinatorial optimization. It was first observed in 1983. Today this algorithm is popular among practitioners because of its simplicity, flexibility and efficiency, as well as among theorists, since the algorithm cannot analytically investigate its properties and prove asymptotic convergence.

Simulated annealing refers to a class of local search algorithms threshold. At each step of the algorithm for this decision in its neighborhood chosen a solution and if the difference in the objective function between the new and the current decision does not exceed the predetermined threshold, the new solution replaces the current one. Otherwise, select the new adjacent solutions. The general scheme of threshold algorithm can be represented as follows.

E. Taboo search

Taboo search is a meta-heuristic algorithm, local search is to prevent it from falling into the trap of premature local optimum by prohibiting the movement of those who are forced to return to previous decisions and cyclic operation. Taboo search begins with the initial decision. At each iteration margin generated solutions and best for this neighborhood gets a new solution. Certain attributes stored in previous decisions taboo-list, which is updated at the end of each iteration. Choosing the best solution going around so that it does not take any of the prohibited attributes. Most acceptable solution is now being updated when new and better solution to the current limit. The procedure continues until the fail any of the two criteria breakpoint, which is the maximum number of iterations performed and the maximum number of iterations during which the current solution is not improving.

III. THE FEASIBILITY OF THE RESEARCH

An interesting example is the guild task scheduling theory routing generalizing guild tasks commonly known metric traveling salesman problem. Performances such problems appeared independently in considering problems arising both in manufacturing and in the service industry. Most tasks scheduling theory is NP-difficult.

Sorting algorithms exponential type require considerable computational cost even when solving examples of average dimensions. Therefore, one of the important areas of research is accelerating algorithms for NP-hardness. Currently, this area has gained a huge number of supporters among researchers engaged in computer mathematics. Here are the main reasons for the popularity:

- there are many optimization problems that require solutions, and most NP-complex. For many of them not necessarily find the exact solution, but rather to build about, but in a short time, the more successfully cope with approximate algorithms;
- in practice, most optimization problems are very complex, because it includes many additional restrictions that do not allow to find a satisfactory approximate solution. But usually approximate algorithms for simpler versions of the same tasks suggest new ideas for heuristics, which then successfully used for practical tasks;
- in terms of exact solutions of NP-almost all challenges are equivalent to each other. Construction of approximate algorithms or proved the impossibility of constructing differences in the classes P and NP makes it possible to compare th NP-complex tasks by how they can be approximated.

Thus, we see that the current algorithms can be improved at the expense of speeding.

IV. FORMULATION OF THE PROBLEM

The problem purpose partially ordered work to handle their various capacities on machines that contain certain periods of downtime if downtime to minimize the time of execution of works. Suppose we have a set of works

$J = \{J_i\}, i = \overline{1,n}$ i and set of cars $M = \{M_j\}, j = \overline{1,m}$. Time for each job on each machine defined matrix $(x_{ij})_{n\times m}$. Also considered a given set period of inactivity $H = \{H_l\}, l = \overline{1,k}$, along with three vectors $(\beta_1,\ldots,\beta_k)$, $(r_1,\ldots,r_k)$, $(\tau_1,\ldots,\tau_k)$, which determine the length of downtime, the number of machines to which he belongs, and the time it starts, respectively. The problem can be represented by two vectors — $(s_1,\ldots,s_n)$, $(z_1,\ldots,z_n)$, which specifies the number of machines on which work is scheduled and the start of its implementation. Based on the above notation, a formal model of the problem can be represented as follows:

$$(x_{ij}) > 0, i = \overline{1,n},\ j = \overline{1,m},$$
$$\beta_l > 0, l = \overline{1,k},$$
$$s_i \in \{1,\ldots,m\}, i = \overline{1,n},$$
$$r_l \in \{1,\ldots,m\}, l = \overline{1,k},$$
$$z_i > 0, i = \overline{1,n},$$
$$\tau_l > 0, l = \overline{1,k},$$

For all $i \in \{1,\ldots,n\}, \nu = \{1,\ldots,n\}$ such that $s_i = s_\nu, z_i < z_\nu$:

$$z_i + x_{is_i} \le z_\nu \quad (1)$$

For all $i \in \{1,\ldots,n\}, \nu = \{1,\ldots,k\}$ such that $s_i = r_\nu, z_i < \tau_\nu$:

$$z_i + x_{is_i} \le \tau_\nu \quad (2)$$

For all $i \in \{1,\ldots,n\}, \nu = \{1,\ldots,k\}$ such that $s_i = r_\nu, \tau_\nu < z_i$:

$$\tau_\nu + \beta_\nu \le z_i \quad (3)$$

$$\max_{i=1,n}(z_i + x_{is_i}) \to \min \quad (4)$$

Limit (1) sets no crossing of the planned works on the same machine, constraints (2) and (3) - not crossing works and downtime. Limitation (4) - objective function that provides minimizing completion time of the most recent (v. BC. Guillotine cutting).

V. ALGORITHM TO OBTAIN AN INITIAL ACCEPTABLE SOLUTION

Step 1. Choose a car with the most power.
Step 2. Go to Step 4.
Step 3. CHOOSE car with less power.
Step 4. CHOOSE task with the greatest of less than or equal intervals to stop the machine.
Step 5. delete the selected task from the list available.
Step 6. Set the selected task in the processing machine.
Step 7. If the current machine is not the least power, then go to Step 3.
Step 8 select the first car
Step 9. Go to Step 11.
Step 10. Step the Next machine
Step 11. CHOOSE task with the greatest of less than or equal intervals to stop the machine.
Step 12. If the problem is not selected TO CHOOSE next interval after stopping the car OTHERWISE go to step 14.
Step 13. Go to Step 11.
Step 14. Set the time the machine flat end time of the end processing tasks.
Step 15. If the current machine is not the last, then go to Step 10.
Step 16. CHOOSE car with the lowest end of time.
Step 17. delete the selected task for this machine from a list.
Step 18. Set the selected task in handling this machine.
Step 19. If the list of available tasks does not blank, then go to step 8.

CONCLUSIONS

As the previous review, all known methods for scheduling the task to be performed with NP-complexity. Given the current trends in computing, it should focus on reducing the complexity of existing algorithms, which will accelerate their work and productivity software that will use them. Were considered well-known methods for solving the above problem and algorithm to obtain an initial acceptable solution.

REFERENCES:

[1] Hulyanytskyy L, Turin V. “Mathematical model of a class of independent tasks scheduling machines”, *Computer mathematics. Kyiv Institute of Cybernetics, vol. 2*, 2012, pp. 1-6.
[2] S. Kolliopoulos, G. Steiner. “Approximation algorithms for minimitation the total weighted tardiness on single machine”, 2006, pp.1-10
[3] A. Lazarev, E Rashidovich “Theory of schedules”, Moscow, 2011, pp. 222.

Recognition of fragments of standard images at low light level and the presence of additive impulsive noise

V. Avramenko, K. Salnik
e-mail: avr@sumdu.edu.ua, kristina.sal97@gmail.com

***Abstract* – On the basis of integral disproportion function of the first-order the algorithm recognizing fragments of standards is created. It works in low light image that is analyzed and the presence of additive impulse noise. This algorithm permits to fin an appropriate pixel in one of several standards for each pixel of the image.**

***Keywords:* disproportion function, fragments of standards, image recognition, low lighting, additive impulse noise, bipolar impulse noise .**

I.INTRODUCTION

The recognition of the fragments of standard photo and video images on the image, that is analyzed, often is occur in practice. For example, some standard objects or the parts of them may be present on the terrain. Some objects may be imposed on other one or to be closer to each other. In addition, they may be different scale images or to be shifted and rotated compared to the standards. Besides of they may be distorted due to geometric transformations. For digital camera a camcorder the images are represented as two-dimensional arrays of pixels. Each pixel has its own color value. These are transparency and intensity of red, green and blue components. Each of them can vary from 0 to 255. There are many papers devoted problem of image recognition. In particular, a very effective tool is to normalize distorted images [1]. Pattern recognition based on fuzzy neural classifier provided the facility that is recognized correctly positioned relative coordinate axes, is proposed in [2].

However, the problem of recognizing fragments of standards in low light image that must be analysed, is actual. Basically, it is recommended to use a more sensitive equipment or devices that operate in the infrared spectrum. However, even in the daytime lighting levels can deteriorate rapidly. In addition, obtaining video images may occur when there are the smoke, the fog, or the sediments. All this leads to a weakening of the intensities of the components that determine the color of the pixel in the digital image. It is possible that may arise the conditions under which these components will be reduced differently. Compared to the brightness of red, green and blue components of the standards, these components of the images that are analyzed, will be less. The attenuation may be equal for all components, in case there is reducing lighting, and it is different, if the image is fixed through a medium which absorbs differently frequencies of the visible spectrum.

In any case, the problem is not so much in getting the intensities of brightness components of pixels. It's necessary to recognize the fragment of standard to which every pixel of image corresponds. In fact, there is a need for recognition in a conditions, where each component color image pixel intensity is proportional to the intensity of the standard, but the factor of proportionality is unknown. Its value can't be determined simply by dividing the intensities for pixels that are randomly taken from image and from standard one. It can be explained because it isn't known the standard for which the pixel selected from image corresponds. Thus, even in circumstances where an image that is analyzed, has the same size, isn't shifted and isn't returned relatively standard one, it isn't simply to identify the fragment of this standard.

In fact, it's necessary to check the proportional relation between the color of every pixel of image and the pixel color for each of standards.

Another problem is solving the above problem in the worse conditions, when there is reducing lighting and a noise is superimposed on a video signal.

Getting and video signal transmission usually occurs in the presence additive noise as well as multiplicative one. Often there is an additive bipolar impulse noise.

This signal has a peculiarity. It has the impulses often are more than video signal amplitude. In this case the pixels, which got the pulses of noise become either white or black [3].

The cause of bipolar pulse noise may be a result of electric welding, electrical discharges, switching processes in electrical circuits etc. Thus, we must first recognize a moment when the noise disappeared and to recognize the fragment of standard before the noise will arise again.

Often the impulse noise appears and disappears at random time. That is, the noise is a random process. In practice, the statistical characteristics of this process are unknown. This doesn't allow to implement it's effective filtering.

In principle, recognizing a fragment of standard image can be in those intervals when the noise disappears, but these intervals must also be recognized. The solving of this problem is complicated because according conditions it's unknown which pixel of image corresponds to which pixel of standard at current time.

Moreover, this standard is unknown also. In [4] this problem is considered for the case where the obstacle is described by smooth function. However, the impulse noise mostly often is not smooth.From the outset let's consider solving the problem when there is no noise.

II. FORMULATION OF THE PROBLEM IF A NOISE IS ABSENT

There are m reference images (standards), represented by matrices of pixels.

After scan there are the arrays of red $R_k[q]$, green $G_k[q]$ and blue $B_k[q]$ brightnesses for every pixel and for every standard, where k = 1,2, ... m - the order number of the standard;q = i * w + j - the order pixel number;w - the number of pixels in one line;j - the order pixel number in a row;i - line number $0 \leqslant i \leqslant h$; h - the number of rows.

Also there are the arrays of brightness's r[q], g[q] and b[q] respectively red, green and blue components of the color pixels of image, that must be analyzed. It is obtained in low light. This image consists of fragments of standards that must be recognized. In the case where a pixel in the image corresponds to the pixel k- th standard, their brightness are proportional:

$$r[q]=k_rR_k[q]; \quad (1)$$

$$g[q]=k_gG_k[q]; \quad (2)$$

$$b[q]=k_bB_k[q], \quad (3)$$

where k_r, k_g, k_b - the coefficients attenuation brightness's.

In general, these factors may be different depending on the medium through which light passes. In addition, they might accidentally change over time, for example, by passing clouds or of smoke. Thus, to solve problem it's necessary to find for every pixel of image the appropriate pixel of standard. In this case will be a proportional relation between brightness's of these pixels at least for one component. But it must be executed in case, when the coefficients of attenuations for brightness's are unknown and their values are random. So it's necessary to use the method is invariant to the coefficient of proportionality.

III. SOLVING THE PROBLEM IF THE NOISE IS ABSENT

The disproportion function by derivative of first range for numeric functions, that are defined as parametric , corresponds for all conditions, that were showed above [5].

This disproportion function of y (t) with respect to x(t) is described by the expression:

$$z(t) = @d^{(1)}{}_{x(t)}\, y(t) = \frac{y(t)}{x(t)} - \frac{dy/dt}{dx/dt}, \quad (4)$$

where @ - character that defines the computing disproportion;

d - means derivative.

There is an order of disproportion in round brackets.

Reading: «at d one of y (t) with respect to x (t)».

In case if

$$y(t)=kx(t), \quad (5)$$

where k - constant factor,

disproportion (4) is equal to zero regardless of the value of the coefficient k in (5).

Note that the brightnesses (1), (2), (3), as well as the brightnesses of standards are discrete.

That's why an integral disproportion function of first order [6] is proposed instead of disproportion (4) .

The integrtal disproportion function of r[q] with respect to $R_k[q]$ has the form:

$$I[q]=@I^{(1)}_{R_k[q]}r[q] = \frac{r[q-1]+r[q+1]}{R_k[q-1]+R_k[q+1]} - \frac{r[q]}{R_k[q]} \quad (6)$$

It is proposed calculate the disproportion (6) for the color components of each image pixel with respect to color components of each pixel for all standards. Further driven algorithm for solving the problem for example when analyzed red brightness .

1 Read brightness $R_k[q]$ of red component for pixels of standards: (k = 1,2, ... m), (q = 1,2, ... hw);

2 Read brightness r[q] of red component for image pixels: q=1,2,…hw;

3 Set the standard k=1;

4 Calculate disproportion (6) of r [q] with respect to R_k [q] and to store their values (q = 1,2, ... hw);

5 k = k + 1;

6 If k <= m, go to p 4. Otherwise – go to p 7;

7 Show the results for each standard.

This can be done in different ways. You can simply print the values of disproportions and can reproduce images on the screen, which will display the pixels of the standards for which disproportion (6) is zero. It is not excluded that some standards may have identical fragments. The algorithm tested on test case in the processing of all 3 components. As standards two photos (Fig.1) are used.

Figure 1. The standards

One of them are the trees over a river. Another – a cat is sleeping on the computer table. Further the image for recognition is showed (Fig.2). The part of this image is a fragment of computer table and further is a part of a cat. The next part is a foto of trees. Due to a low level of

lighting this image is almost black. In these cases, it's difficult to recognize these fragments. So the algorithm that is proposed was used. As a result, the disproportions (6) have received zero values for the corresponding pixels of corresponding standards. On Fig.2 there is an image for showing the results. It was built with using disproportion values. On this image the fragments of both etalons are recognized. So the test passed successfully.

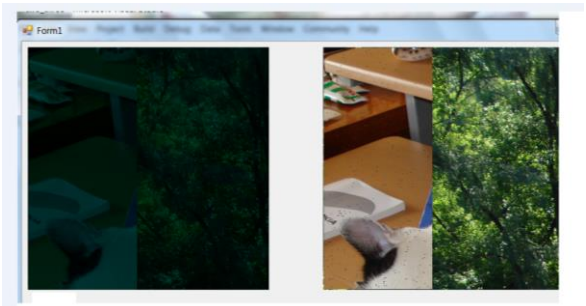


Figure 2 – Image for recognition (black), image of results

IV. RECOGNITION AT LOW-LIGHT IMAGES AND WITH PRESENCE OF IMPULSE ADDITIVE NOISE

Let's consider this problem, for example, for red component r [q] color of q - th pixel image, that is analyzed,only.

When the brightness's of q-th image pixel are fixed, the noise η (t) is added. Thus, the noise impact is tied with pixel. The noise may be represented by η[q], q = 1,2, ... wh.

So in the presence of additive noise the brightness signal of red component has the form:

$$y[q]=r[q] + \eta[q], \quad (7)$$

The next expression was received after substituting expression (1) to (7) :

$$y[q]=k_r R_k[q] + \eta[q], \quad (8)$$

Under conditions of the problem k_r - is unknown.

The standard k, which corresponds to a pixel of image, is unknown also.

As in the previous case, for each k = 1,2, ... m the disproportion (4) of the signal y(t) (8) with respect to $R_k[q]$ is calculated.

$$z_k[q]=\frac{k_r R_k[q] + \eta[q]}{R_k[q]} - \frac{k_r R'_k[q] + \eta'[q]}{R'_k[q]} = @d^{(1)}_{R_k[q]}\ \eta[q]. \quad (9)$$

At a time when the noise disappears the value of disproportion (9) is equal to zero. Thus, if zk[q] = 0, it means no noise and the fact that the image pixel corresponds to the pixel of k - th standard.If for all standards $z_k[q] \neq 0$, it may indicate the presence of noise. Also may be a case where a pixel image does not correspond to any standard.However, the disproportion (9) can be calculated only when the noise is smooth and has a first derivative. In general, the impulsive noise can't be differentiate. Also, again, we must remember that the brightnesses are discrete. Therefore, in this case, the integral disproportion of first-order of signal y[q] with respect to R_k [q] is proposed also:

$$I[q]=@I^{(1)}_{R_k[q]}y[q] = \frac{y[q-1]+y[q+1]}{R_k[q-1]+ R_k[q+1]} - \frac{y[q]}{R_k[q]} \quad (10)$$

Thus, in the case of additive impulsive noise the integral disproportion (10) should be calculated and it must be compared with zero, as was done with the disproportion (9). So, with the presence of additive noise the algorithm given above is used. But now there is recognition only in the intervals when there is no noise.Obviously, the algorithm requires significant computing resources, but at the same time, it allows parallelization of computing. For example, you can simultaneously work with multiple standards.It should also be borne in mind that the implementation of the algorithm allows only get information for decision-making system. This system relates the fragment of an image with a standard.

V. CONCLUSIONS

The algorithm for analyzing of image to recognize on it some fragments of standards is proposed. The lighting of image may be low. In addition, the impulsive noise may be added to the video signal of image. For this purpose the integral disproportion function of first order of brightnesses color components of each pixel image with respect to brightnesses of each pixel standard is used. The algorithm can be used in the system of decision making at analyzing photos and video images.

REFERENCES

[1] Putjatin E.P. Normalizacija i raspoznavanie izobrazhenij
[2] http://sumschool.sumdu.edu.ua/is-02/rus/lectures/pytyatin/pytyatin.htm
[3] 2 Poltorak V.P., Dorogoj Ja.Ju.Sistema raspoznavanija obrazov na baze nechetkogo nejronnogo klassifikatora. http://aaecs.org/poltorak-vp-dorogoi-yayu-sistema-raspoznavaniya-obrazov-na-baze-nechetkogo-neironnogo-klassifikatora.html.
[4] 3 Krasil'nikov N.N. Cifrovaja obrabotka 2-D i 3-D izobrazhenij : uchebn. posobie.- SPb: BHV – Peterburg, 2011.- 608s.
[5] 4 Avramenko V.V. , Slepushko N.Ju. Raspoznavanie jetalonnyh signalov pri nepolnoj informacii o harakteristikah pomeh//Visnik SumDU. Serija Tehnichni nauki. №4 2009, s.13-19.
[6] 5 Avramenko V.V. Harakteristiki neproporcional'nostej i ih primenenie pri reshenii zadach diagnostiki/ V.V. Avramenko // Vestnik SumGU. – 2000. – №16.
[7] 6 Karpenko A.P. Integral'nye harakteristiki neproporcional'nosti chislovyh funkcij i ih primenenie v diagnostike // Vestnik Sumskogo gos. Un-ta.-2000.-№16.- s.20-25.

Gait recognition on base of representation in spatio-temporal area

Mihail Babiy
Sumy State University, mbabiy@id.sumdu.edu.ua

***Abstract* – New method of automatic gait recognition from video is proposed. The method works with a sequence of silhouettes derived from the video after background subtraction, decreasing shadows and noise. Two-dimensional silhouette shape is converted into one-dimensional signal presenting distance from center of gravity to outline of this silhouette. A set of signals extracted from a sequence of silhouettes forms a two-dimensional picture.**

The features extraction is performed using Gabor wavelets. Testing the method on the samples of CASIA gait database showed high recognition accuracy.

***Keywords* – gait, recognition, Gabor wavelets, CASIA database.**

I. INTRODUCTION

Automated person identification is important research problem in the area of computer vision. Traditional face recognition technology is applicable only to frontal or nearly frontal faces. In addition, the recognition capabilities are very limited at night, when only silhouette of a person can be observed. In these cases, gait recognition can be a promising method.

Gait is a spatio-temporal process that typifies the motion characteristics of an individual. As biometric characteristic, gait has next advantages: (1) invisibility for the object under surveillance, (2) no need to improve the quality of the image, (3) the complexity of disguise, (4) the possibility of identifying at a significant distance from the place of observation, when other attributes are still indistinguishable.

There are two main ways to represent gait in recognition problems: model-based approach and model-free approach. In the first case, static and dynamic body parameters are determined in modeling the movement of the arms, legs and other body components. However, model-based approach is sensitive to the quality of the gait image sequences. Another disadvantage is large amount of computation.

Model-free approach focus on shape of silhouette. This approach is insensitive to the quality of silhouettes. It requires less computation and is more often used in practice.

In the framework of this approach, some authors apply the hidden markov models (HMM). Iwamoto et al. [1] regards the walking operation as a periodic signal. A period (gait cycle) is defined as the time of heel strike for the same leg. Person's outline information in each frame is used as the feature parameter. The feature vector is compressed by principle component analysis (PCA). Training and identification processes are performed using continuous HMM

In [2] the discrete HMM is taken into consideration for recognition process. Vector quantization and clustering are used to convert continuous sequence of gait signals into a discrete sequence of symbols for discrete HMM. A general drawback of this approach is ambiguity in specifying the number of clusters and the number of hidden states.

Topological approach is used in [3] and [4]. 3D digital picture is obtained by gluing silhouettes through their gravity centers. A border simplicial complex associated with picture is constructed. Authors define eight directions to obtain eight filtrated simplicial complex. According to each filtration a persistence barcodes are calculated. Using barcodes the similarity value for two gait subsequences can be computed. A triangulation is supported by existing software, but number of required triangles is very large and the computational analysis by topological approach is correspondingly slow.

II. CREATING THE GAIT IMAGES

The proposed algorithm includes next phases: silhouette estimation, transformation of gait sequence to gait picture, feature extraction using wavelets, near-neighbor gait recognition "Fig. 1".

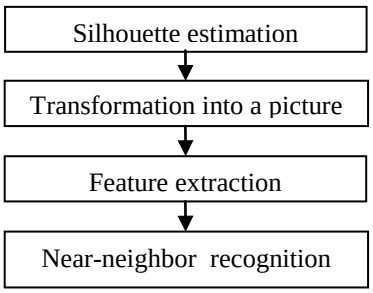


Figure 1. Flowchart of algorithm

Background modeling uses median method. This method incorporates each foreground pixel so that each occurrence of foreground objects leads to corruption of

background image. To obtain the silhouette, background subtraction and binarization is performed.

At the next stage, the gravity center of silhouette is determined. By choosing the gravity center as reference origin, we unwrap the outer contour clockwise. For this purpose, we find distances from the gravity center to the contour points lying on the rays emerging from this center "Fig. 2".

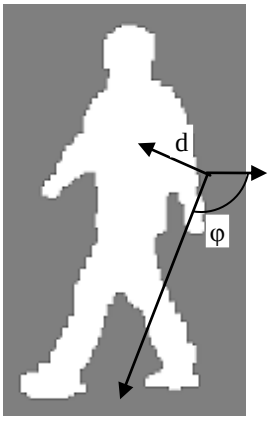


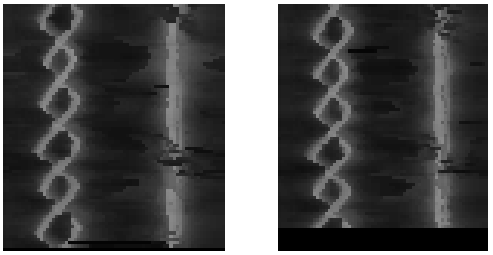
Figure 2. Determining distances from the gravity center to the outer contour of silhouette

In our research, the angle between neighboring rays $\Delta\varphi = \pi / 36$. As a result, the distance signal $D = \{d_1, d_2, \ldots, d_{72}\}$ is obtained. This signal indirectly represents the initial 2D silhouette shape into 1D space.The algorithm for determining a distance is as follows. Starting from the gravity center, we move along the ray with increment equal to one. The following actions are performed at each step. Using the current coordinates, coordinates of the neighboring pixel $\{x_p, y_p\}$ are calculated as $x_p = \text{floor}(x)$, $y_p = \text{floor}(y)$. If the brightness of the pixel $\{x_p, y_p\}$ becomes less than the threshold value, the distance is taken equal to the number of steps. Otherwise the movement along the ray continues. For noisy silhouettes, other algorithm can be used. The motion is performed to the image boundary. The distance is defined as the number of steps, by which the brightness of the current pixel is greater than the threshold value.Wang and Tan [5] indirectly use obtained signals for person identification. They apply PCA training to represent the original gait features from high-dimensional measurement space to a low-dimensional eigenspace. Nonzero eigenvalues and the associated eigenvectors based on singular vector decomposition (SVD) are calculated. After that an original distance signal can be projected into a point in the eigenspace.

Other approach is used in this article. For each frame in the time sequence of the silhouettes, a corresponding 1D signal is constructed. We introduce the coordinate *s*, representing the numbers of the border points in which the distance signal is measured. By grouping the silhouettes at sequenced times *t*, the gait can be represented as a graph of the signal *D* in the spatio-temporal area (*s*, *t*). We convert this graph into the normalized grayscale image *I*(*s*, *t*) by formula

$$I(s,t) = \frac{255D(s,t)}{d_{\max}},$$

where $d_{\max}$ is the maximal value of *D(s, t)*.
In the following text, instead of the notations (*s*, *t*), we will use the more habitual symbols (*x*, *y*) to address the image points. We will call these images "gait images". Examples of images are shown in "Fig. 3".



(a) (b)

Figure 3. Examples of gait images from CASIA database: (a) gait fyc/00_1; (b) gait fyc/00_3.

III. FEATURE VECTOR EXTRACTION

Gait images are converted to standard size 128*128 pixels. Feature vector is constructed from the responses of the gait image to Gabor filters.

Two-dimensional Cabor wavelet has the shape of a plane wave bounded in amplitude by a Gaussian function. Each wavelet ψ_j from a given set of wavelets is determined by its wave vector $\vec{k}_j$ [6]:

$$\psi_j(\vec{r}) = \frac{k_j^2}{\sigma^2}\exp\left(-\frac{k_j^2 r^2}{2\sigma^2}\right)\left[\exp(i\vec{k}_j\vec{r}) - \exp\left(-\frac{\sigma^2}{2}\right)\right], \quad (1)$$

where $\vec{r} = \vec{r}(x, y)$. We define the vector $\vec{k}_j$ for five different frequencies with indices $p = 0, \ldots, 4$ and eight different orientations with indices $q = 0, \ldots, 7$:

$$\vec{k}_j = (k_{jx}, k_{jy}) = (k_p \cos\varphi_q, k_p \sin\varphi_q);$$

$$k_p = 2^{-\frac{p+2}{2}}\pi; \quad \varphi_q = q\frac{\pi}{8}; \quad j = q + 8p.$$

Our image is given by array $I(\vec{r})$ of pixel brightness in the points $\vec{r}(x, y)$. Wavelet transform of image can be represented as convolution $R_j(\vec{r}_0)$ of image with *j*-th wavelet from the set of Gabor wavelets:

$$R_j(\vec{r}_0) = \int I(\vec{r})\,\psi_j(\vec{r} - \vec{r}_0)\,dxdy. \quad (2)$$

The feature vector for the image will be constructed from the values of R_j at lattice nodes with a step of 8 pixels horizontally and vertically. This step is quite

acceptable for images with a side length of about a hundred pixels.To enter the metric, we agree to consider the distance between neighboring pixels equal to one. Then, if we choose $\sigma = 2\pi$, then for the wavelet with the index $p = 2$ the standard deviation of the Gaussian function is $\sigma / k_p = 8$, which is equal to the distance to the next lattice node. Accordingly, for $p = 0$, the deviation will be two times smaller, and for $p = 4$ it will be twice as large. Direct calculation of the convolution

$$R(\vec{r}) = I(\vec{r}) * \psi(\vec{r}) \quad (3)$$

is quite laborious. We apply the discrete Fourier transform F to both parts of (3). As a result, the convolution operation is transformed into multiplication of spectra

$$F[R(\vec{r})] = F[I(\vec{r})]\, F[\psi(\vec{r})] \quad (4)$$

Performing the inverse Fourier transform, we obtain

$$R(\vec{r}) = F^{-1}\{F[I(\vec{r})]\, F[\psi(\vec{r})]\} \quad (5)$$

The discrete Fourier transform for the two-dimensional case has the form

$$f_{k_x k_y} = \sum_{n_x=0}^{N_x-1} \sum_{n_y=0}^{N_y-1} x_{n_x n_y} \times$$

$$\times \exp\left(-\frac{2\pi i}{N_x} k_x n_x\right) \exp\left(-\frac{2\pi i}{N_y} k_y n_y\right)$$

For its implementation, it is convenient to use the Fast Fourier Transform (FFT) algorithm. The processing of graphic images is carried out using the non-commercial OpenCV library. It includes a function *cvDFT*, which implements the algorithm for direct and inverse FFT, including for two-dimensional input data. A trait of the FFT is the reduction of the problem for N numbers to the problem for $N1 = N / m$, where m is the divisor of the number N. In the implementation of *cvDFT*, the optimal size of the image side is the product of integer powers of numbers 2, 3 and 5. Accordingly, the size of the side of the image is equal to such a product. To reduce computational complexity, we confine ourselves to the real (even) component of the Gabor wavelet. In this case, the amount of input data is reduced by half, and for complex output data, a special packed format is used, which also reduces the output data by half.To relate an object to a particular recognition class, the nearest-neighbor method is used.

IV. Experimental Results

In accordance with the described algorithm, a program for gait recognition is developed. The program is written in C++ for the Visual Studio environment. Computer image processing is performed using the additional OpenCV library compiled for Visual Studio. For the testing of the program, CASIA Gait Database (DatasetA) was taken [7]. DatasetA includes gait samples for 20 persons. Each person has 12 image sequences, 4 sequences for each of the three directions. The lengths of the sequences are not identical due to the different walker's speed. For testing, the first three persons selected in alphabetical order were taken. The format of the image filename in DatasetA is "xxx-mm_n-ttt.png", where xxx – subject id, mm – direction, n – sequence number, ttt – frame number in a sequence. As a measure of closeness, the similarity function was taken. The results of the calculation are presented in the Table I.

TABLE II. Results of Recognition

Test set	Values of the similarity function for the training set		
	fyc/00_1	hy/00_1	ljg/00_1
fyc/00_3 (fyc-00_3-001.png – fyc-00_3-069.png)	0,893	0,874	0,865
hy/00_3 (hy-00_3-001.png – hy-00_3-060.png)	0,817	0,907	0,867
ljg/00_3 (ljg-00_3-001.png – ljg-00_3-069.png)	0,866	0,851	0,869

Thus, all three gaits are correctly recognized.

V. Conclusions

In this paper we propose the new method of automatic gait recognition from video. Two-dimensional silhouette shape is converted into one-dimensional signal. A set of signals extracted from a sequence of silhouettes forms a two-dimensional picture. The features extraction is performed using wavelet analysis. Testing the method on the samples of CASIA gait database showed high recognition accuracy.

References

[1] K. Iwamoto, K, Sonobe, N. Komatsu, "A gait recognition method using hmm," *Proceeding of the SICEAnnual Conference Roy. Soc. London,* August 2003, pp. 1936–1941,

[2] A. Kolawole, A. Akintola, "A novel gait recognition system based on hidden markov models," *Proceedings of the 8th International Symposium on Visual Computing*, 2012.

[3] J. Lamar-Leon, E. Garcia-Reyes, R. Gonzalez-Diaz, "Human gait identification using persistent homology," *CIARP*, *Lecture Notes in Computer Science*, vol. 7441, pp. 244–251, Springer, 2012.

[4] J. Lamar, E. Garcia-Reyes, R. Gonzalez-Diaz, R. Alonso-Baryolo, "An application for gait recognition using persistent homology," Univ. of Seville, Spain, 2013.

[5] L. Wang, T. Tan, "Silhouette analysis-based gait recognition for human identification," *IEEE Trans. on pattern analysis and machine intelligence*, vol. 25, no. 12, pp. 740–741, December 2003.

[6] L. Wiskott, J. M. Fellous, N. Kruger, C. Malsburg, "Face Recognition by Elastic Bunch Graph Matching," *IEEE Trans. on pattern analysis and machine intelligence*, 19(7), pp.775-779, 1997.

[7] CASIA Gait Database (DatasetA) http://www.cbsr.ia.ac.cn/GaitDatasetA-silh.zip

Feature learning for information-extreme classifier

A. Korobov, A. Moskalenko, V. Moskalenko
Sumy State University, v.moskalenko@cs.sumdu.edu.ua

***Abstract* – The feature learning algorithm for information-extreme classifier by clustering of Fast Retina Keypoint binary descriptor, calculated for local features, and usage of spatial pyramid kernel for increasing noise immunity and informativeness of feature representation are considered. Proposed a method of parameters optimization for feature extractor and decision rules based on multi-level coarse features coding using information criterion and population-based search algorithm.**

***Keywords* – coarse coding; Fast Retina Keypoint; feature extraction; machine learning; classifier; information criterion.**

I. INTRODUCTION

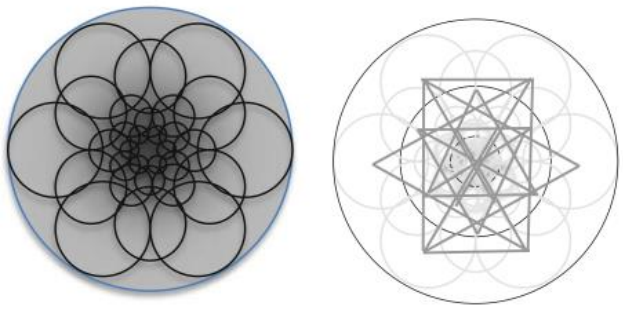
One way to improve throughput and reduce power consumption in autonomous and distributed intelligent systems is the implementation of computationally efficient algorithms. The process of intelligent processing be divided into three stages: collection and data preprocessing, feature extraction and classification or regression analysis. For a detailed analysis of data, common method of filing is the transformation of a two-dimensional matrix, which can be interpreted as an image. That is why in recent years the main direction in evolution of intelligent systems is related to research and development trainable feature extractors for visual data.

Recent experience shows that the most effective approach to the analysis of visual data is to use local features as patch descriptors that describe the variation of the brightness gradient in the neighborhood of the key points of the image [1]. The binary descriptor of key points is faster and simpler. Among them the most effective algorithm in terms of accuracy is FREAK (Fast Retina Keypoint) descriptor, which simulates the behavior of the cells of the eye. However, there is a lack of effective algorithms for descriptor quantization and sparse coding of observations described by binary descriptors. Usage of K-means algorithm of cluster analysis for quantization binary descriptor within byte representation leads to slow convergence, dependance on the choice of initial centroids and low informativeness of obtained quantization [1]. Also high dimension of obtained observations usually leads to the intersection of classes in the feature space, overfitting and the "curse of dimensionality" during parameters regularization for classifiers based on neural networks or support vector [2,3].

A promising approach of system synthesis for intelligent data analysis is using algorithms of self-organize unstructured grids, such as the method of growing neural gas, for quantization of binary descriptors, FREAK, within bit level representation and methods of classification analysis based on the usage of coarse binary encoding floating-point descriptor by multi-level system of thresholds. In this case, transformation of feature space is performed by computationally efficient comparisons. In this case, usage of logarithmic information criterion for optimization that is more effective in condition of restricted amount of training samples owing to the smoothing effect of the logarithmic function is proposed.

II. FEATURE LEARNING

In FREAK the sampling grid shown in Fig. 1, where each circle corresponds to a receptive field and its size represents the standard deviation of the Gaussian kernels applied to the underlying sampling point [1].



a b

Figure 1. Receptive fields (a) and corresponding pairs (b) in FREAK

In an attempt to reduce noise, the smoothed intensities of these receptive fields has been thresholded based on the neighborhood information. In binary FREAK descriptor used thresholded difference between pairs of the receptive field smoothed by corresponding Gaussian kernels. Calculation binary vector performed through one-bit sequence differences:

$$F = \sum_{0 \le a < N} 2^a \, sign\left(I\left(P_a^{r_1}\right) - I\left(P_a^{r_2}\right)\right) \qquad (1)$$

where N – selected descriptor size; I($I(P_a^{r_i})$) – smoothed brightness of i -th field of a - th receptive fields pair. To adapt the algorithm neural gas for cluster analysis binary descriptors on bit level it is proposed some modification. Consider the basic steps of a modified neural gas:

1. Defined two primary nodes a and b in points w_a and w_b , which corresponding to the two first input vector. Nodes a and b are connected by an edge, whose

age is zero. Errors in the nodes a and b similarly taken a zero.

2. Selects the next input vector (signal) x.

3. Determined the nearest and the second, in the sense of closeness to Hamming distance, x nodes (and denote them s_1 and s_2 respectively).

4. Age of all edges, that incidental to node, increase per unit.

5. Counter error node s_1 increase in

$$d(x \oplus w_{S_1}) = \sum_{i=1}^{N} (x_i \oplus w_{s_1}, i).$$

6. s_1 and its topological neighbors (nodes connected to it by an edge) shifting in a direction to him at a distance $\varepsilon_b \cdot d(x \oplus w_{s_1})$ and $\varepsilon_n \cdot d(x \oplus w_n)$, where $0 < \varepsilon_b << 1$ and $0 < \varepsilon_n << \varepsilon_b$ by inverting of not matching bits in randomly selected not matching bytes.

7. If s_1 and s_2 connected by an edge, its age is reset; otherwise between s_1 and s_2 creates a new edge with age equal to zero.

8. All edges in the graph with age greater than $a_{\max}$ deleted and removed nodes with no incident edges (were isolated).

9. If the number of the current iteration a multiple of the value λ (one of the parameters of the algorithm), then carried insertion a new node at the point w_r between the nodes w_q and w_f where q - the node with the highest accumulated error; f - topological neighbor q with the greatest accumulated error. An edge between f and q removed, instead it attached edges between f and r, and also between r and q. At the same time, coordinates of the vector w_r determined by copying vector f (or q) and by inverting the half of distinct with q (or f) bit in randomly selected distinct bytes. Error in nodes f and q reduced by multiplying by a constant $\alpha < 1$; error value of the new node r is initialized by a value that is equal to the value of node error q.

10. The error of all nodes is reduced by multiplying by a constant $\beta < 1$.

11. If the input data is no more or made the maximum taken number of iterations, then stop algorithm.

Encoding observations is proposed to be implemented by constructing a histogram of occurrence of quantized local features using the Average pooling method [3]. At that assign scores of local features to codewords determined by the principle softmin. To take into account of spatial information is proposed to use Spatial Pyramid Kernel [2]:

$$K = I^L + \sum_{l=0}^{L-1} \frac{1}{2^{L-l}} \left(I^l - I^{l+1} \right), \qquad (2)$$

where I^l – histogram at l resolution level, that calculated by applying the intersection function to the histograms of cells at l resolution level.

III. INFORMATION-EXTREME CLASSIFIER

The basic idea of information-extreme classifier lies in a purposeful binary encoding of quantitative features by comparing their values with corresponding thresholds to build a clear partition feature space into classes. Creation of binary training set is carried out according to the rule:

$$x_{m,i,l}^{(j)} = \begin{cases} 1, \ if \ y_{i,\max} \left[1 - \frac{\delta_{i,l}}{\delta_{\max}} \right] \le y_{m,i}^{(j)} \le y_{i,\max}; \\ 0, \ else, \end{cases} \qquad (3)$$

where $y_{i,\max}$ – the maximum value of i-th features in the training sample; δ_i – parameter of l-th receptive field for i-th feature value $\delta_{\max}$ – the maximum value of receptive field.Calculation coordinate values of binary reference vector, Calculation coordinate values of the binary vector x_m reference against which construction takes place in the space in a radial basis Hamming container classes shall according to the rule

$$x_{m,i,l} = \begin{cases} 1, & \text{if} \quad \frac{1}{n_m} \sum_{j=1}^{n_m} x_{m,i,l}^{(j)} > \frac{1}{n} \sum_{k=1}^{M} \sum_{j=1}^{n_k} x_{k,i,l}^{(j)}; \\ 0, & \text{if} \quad \text{else.} \end{cases} \quad \text{i} = \overline{1,\text{N}} \qquad (4)$$

At the same time, iterative process of parameter optimization $\delta_{i,l}$ performed by maximizing the averaged information criterion on set of classes during training of classifier. Its working formula has the form

$$J_m^{(k)} = \frac{K_{1,m}^{(k)} - K_{2,m}^{(k)}}{n_m (\log(2n_m + 10^{-\omega}) + \omega)} * \\ * \log \left(\frac{10^{-\omega} + n_m + [K_{1,m}^{(k)} - K_{2,m}^{(k)}]}{10^{-\omega} + n_m - [K_{1,m}^{(k)} - K_{2,m}^{(k)}]} \right) \qquad (5)$$

where $K_{1,m}^{(k)}$ – the number of events that characterize the membership of sample of class X_m^o to the container of class X_m^o at k step of machine learning; $K_{2,m}^{(k)}$ – the number of events that characterize the membership of

sample of class X_c^o to the container of class X_m^o; ω – constant that regulates the sensitivity of criteria depending on accuracy characteristics.

Admissible (working) domain of the function of information criterion (5) is limited by inequalities $K_{1,m} \geq 0{,}5 \cdot n_m$ and $K_{2,m} < 0{,}5 \cdot n_m$.

IV. RESULTS

The proposed algorithms were tested on a survey data matrix of EMG sensors. Each observation was presented in the format of HD-EMG map with resolution 72x100 [5].

Training sample contains 300 observations per class. Fig. 1a shows the dependence of parameter λ and information criterion, known algorithm neural gas for quantization the byte representation of binary descriptors and fig. 1b shows the dependence of parameter λ and information criterion modified algorithm neural gas for quantization binary descriptors on bit level.

Search maximum of information criterion carried by the algorithm Particle swarm optimization in parameter space $\delta_{i,l}$. The number of agents swarm was equal to 30, no particle acceleration [6 ,7].

Analysis of Fig. 2 and Fig. 3 shows that the modified algorithm of neural gas provides a maximum value of information criteria at higher values of the parameter λ, which corresponding to a smaller number of clusters.

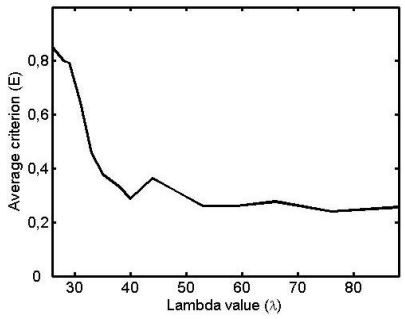


Figure 2. Results of machine learning with classic algorithm neural gas

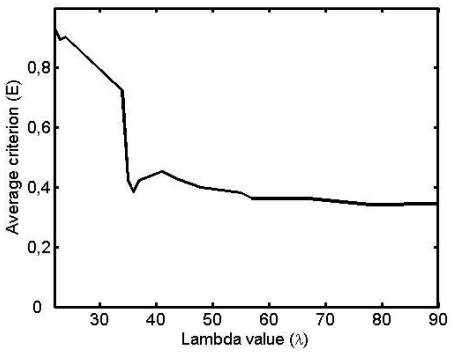


Figure 3. Results of machine learning with modified

At the same time quantization of descriptors on the bit level unlike quantization of descriptor on the byte level allowed to obtain highly accurate decision rules that are close to unmistakable.

V. CONCLUSIONS

So features learning algorithm for information-extreme classifier is the consistent implementation of transformation: construction of two-dimensional maps of signals, calculating local features, quantization of descriptors, coding observations using Spatial Pyramid Kernel and optimization of quantization parameters during the learning process of information-extreme classifier. Moreover, the proposed modification of the neural gas for quantization of binary descriptors on bit level can increase the accuracy of decision rules at a lower computational cost.

REFERENCES:

[1] Hilario Gomez C., Medathati K., Kornprobst P., Murino V., Sona D. "Improving FREAK Descriptor for Image Classification," Lecture Notes in Computer Science, vol. 9163, pp. 14–23, June 2015 [Digests 10th International Conference, ICVS, Copenhagen, Denmark, July 6-9, 2015]

[2] Qiong Yao, Xiang Xu., "Freak Descriptor With Spatial Pyramid Kernel For Scene Categorization," Inte. Conference on Mechatronics, Electronic, Industrial and Control Engineering Published by Atlantis Press, pp. 126–129, April 2015.

[3] S. Gao, L. Duan, and I. W. Tsang, "DEFEATnet–A Deep Conventional Image Representation for Image Classification, " IEEE Transactions on Circuits and Systems for Video Technology, vol. 26, no. 3, pp. 494-505, March 2016

[4] V. Moskalenko, A. Korobov, "Information-extreme algorithm of the system for recognition of objects on the terrain with optimization parameter feature extractor," Radio Electronics, Computer Science, Control., in press. (in Ukrainian)

[5] I. S. Jacobs and C. P. Bean, "High-density surface EMG maps from upper-arm and forearm," Journal of NeuroEngineering and Rehabilitation, vol. 60, issue 2, pp 2–17, December 2012.

[6] A. S Dovbysh, A. S Moskalenko, V. V. Moskalenko, I. V. Shelehov, "Designing algorithms for optimization of parameters of functioning of intelligent system for radionuclide myocardial diagnostics," Eastern-European Journal of Enterprise Technologies,vol. 3, no. 3/9 (81), pp. 11-18, May 2016.

[7] V. Moskalenko, S. Pimonenko, "Optimizing the parameters of functioning of the system of management of data center IT infrastructure," Eastern-European Journal of Enterprise Technologies, vol. 5, no. 3 (83), pp. 21-29, September 2016.

Detecting bivariate outliers on the basis of normalizing transformations for non-Gaussian data

S. Prykhodko, N. Prykhodko, L. Makarova, O. Kudin, T. Smykodub, A. Prykhodko.
Admiral Makarov National University of Shipbuilding, sergiy.prykhodko@nuos.edu.ua, http://ikitn.nuos.edu.ua/

***Abstract* – The statistical technique for detecting outliers in bivariate non-Gaussian data on the basis of normalizing transformations, prediction ellipse and a test statistic (TS) for the Mahalanobis squared distance (MSD), which has an approximate *F* distribution, is proposed. Application of the technique is considered for detecting outliers in two bivariate non-Gaussian data sets: the first, actual effort (hours) and size (adjusted function points) from 145 maintenance and development projects, the second, effort (hours) and mass (tonnes) of designed the section of the ship from 188 designs of sections.**

***Keywords* – outlier; normalizing transformation; bivariate non-Gaussian data; Mahalanobis squared distance; F distribution; prediction ellipse.**

I. INTRODUCTION

An important step in data processing is the outlier detection. Today the problem of outlier detection in a bivariate data set is solved with different methods including statistical [1, 2]. However, well-known statistical methods (for example, bivariate outlier detection based on a prediction ellipse or a test statistic (TS) for the Mahalanobis squared distance (MSD), which has an approximate the *F* distribution) are used to detect outliers in a data set under the assumption that the data is generated by a bivariate Gaussian distribution. And this assumption is valid only in particular cases. In [3] and [4] statistical outlier detection techniques for multivariate non-Gaussian data on the basis of normalizing transformations and MSD, which has an approximate the Chi-Square distribution and the *F* distribution respectively, were proposed. We propose a statistical outlier detection technique for bivariate non-Gaussian data on the basis of normalizing transformations, prediction ellipse and TS for MSD, which has an approximate *F* distribution. The technique consists of two steps. In the first step, bivariate non-Gaussian data is normalized using a bivariate normalizing transformation. In the second step, MSD, prediction ellipse and TS for MSD are calculated and compared with a quantile of the *F* distribution. The data values for which a value of TS for MSD is greater than the quantile of the *F* distribution are considered as outliers and these values are cut off. Two steps should be repeated for the data after outlier cutoff until all values of TS for MSD will be less than or equal to the quantile of the *F* distribution.

II. THE STATISTICAL TECHNIQUE

The outlier detection technique for bivariate non-Gaussian data is based on normalizing transformations, a prediction ellipse and a test statistic for MSD, which has an approximate *F* distribution. Consider bijective bivariate normalizing transformation of non-Gaussian random vector $\mathbf{X} = \{X_1, X_2\}^T$ to Gaussian random vector $\mathbf{Z} = \{Z_1, Z_2\}^T$ is given by

$$\mathbf{Z} = \psi(\mathbf{X}). \qquad (\square\square\square)$$

The values of the sample observations or bivariate data points $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_N$ are normalized using the transformation (1).

The Mahalanobis squared distance for each bivariate data point *i*, $i = 1,2,\ldots,N$, is denoted by d_i^2 and given by

$$d_i^2 = (\mathbf{Z}_i - \overline{\mathbf{Z}})^T S_N^{-1} (\mathbf{Z}_i - \overline{\mathbf{Z}}), \qquad \square(2)$$

where $\overline{\mathbf{Z}}$ is the sample mean vector and S_N is the sample correlation matrix

$$S_N = \frac{1}{N} \sum_{i=1}^{N} (\mathbf{Z}_i - \overline{\mathbf{Z}})(\mathbf{Z}_i - \overline{\mathbf{Z}})^T . \qquad \square(3)$$

A test statistic for d_i^2 can be created as follows [5]

$$N(N-2) d_i^2 / 2(N^2 - 1), \qquad \square(4)$$

which has an approximate *F* distribution with 2 and $N-2$ degrees of freedom.

The equation for the prediction ellipse is defined by [6].

$$(\mathbf{Z} - \overline{\mathbf{Z}})^T S^{-1} (\mathbf{Z} - \overline{\mathbf{Z}}) = \frac{2(N^2 - 1)}{N(N-2)} F_{2,N-2,\alpha}, \qquad (5)$$

where $F_{2,N-2,\alpha}$ is a quantile of the *F* distribution; α is significance level. We take α as 0.05.

A test statistic for MSD (4) is compared with $F_{2,N-2,\alpha}$. The data values for which a value of TS (4) is greater than the quantile of the *F* distribution are considered as outliers and these values are cut off. After outlier cutoff the reduced number of bivariate data points are normalized using the transformation (1) again until all values of TS (4) will be less than or equal to the quantile of the *F* distribution.

III. BIVARIATE NORMALIZING TRANSFORMATIONS

Some transformations have been proposed for normalizing multivariate non-Gaussian data, such as, transformation on the basis of the decimal logarithm, the Box-Cox transformation, the Johnson translation system and others. However, only a few normalizing transformations are bijective. Such bijective transformation is the transformation of S_U family of the Johnson translation system. The Johnson normalizing translation is given by [7]

$$\mathbf{Z} = \boldsymbol{\gamma} + \boldsymbol{\eta}\mathbf{h}\left[\boldsymbol{\lambda}^{-1}(\mathbf{X} - \varphi)\right] \sim N_m(0_m, \Sigma), \quad (6)$$

where Σ is the correlation matrix; $\boldsymbol{\gamma}$, $\boldsymbol{\eta}$, φ and $\boldsymbol{\lambda}$ are parameters of the Johnson normalizing translation; $\boldsymbol{\gamma} = (\gamma_1, \gamma_2)^T$; $\boldsymbol{\eta} = diag(\eta_1, \eta_2)$; $\varphi = (\varphi_1, \varphi_2)^T$; $\boldsymbol{\lambda} = diag(\lambda_1, \lambda_2)$; $\mathbf{h}[(y_1, y_2)] = \{h_1(y_1), h_2(y_2)\}^T$; $h_i(\cdot)$ is one of the translation functions

$$h = \begin{cases} \ln(y), & \text{for } S_L \text{ (log normal) family;} \\ \ln[y/(1-y)], & \text{for } S_B \text{ (bounded) family;} \\ \operatorname{Arsh}(y), & \text{for } S_U \text{ (unbounded) family;} \\ y & \text{for } S_N \text{ (normal) family.} \end{cases}$$

Here $y = (x - \varphi)/\lambda$; $\operatorname{Arsh}(y) = \ln\left(y + \sqrt{y^2 + 1}\right)$.

IV. EXAMPLES

We consider the examples of detecting outliers in two bivariate non-Gaussian data sets: the first, actual effort (hours) and size (adjusted function points) from 145 maintenance and development projects [8], the second, effort (hours) and mass (tonnes) of designed the section of the ship from 188 designs of sections.

Table I contains the data from 145 maintenance and development projects [8], MSD and TS for MSD for standardized data sample units, which are [2]

$$Z_{ki} = (X_{ki} - \bar{X}_k)/S_{X_k} \quad k = 1,2; \; i = 1,2,\ldots,145. \quad (7)$$

The last column in Table I reveals that projects 3, 9, 38, 51, 101 and 102 are bivariate outliers, since $F_{2,143;0.05}$=3.06.

TABLE III. TS FOR MSD FOR THE STANDARDIZED DATA

Pro - ject	Size (adjusted function points)	Actual effort (hours)	Z_{1i}	Z_{2i}	d_i^2	TS for MS D
1	101.65	485	-0.28088	-0.74421	0.57	0.28
2	57.12	990	-0.31024	-0.46733	0.22	0.11
3	1010.88	13635	0.31859	2.97504	11.17	5.51
4	45.6	1576	-0.31783	0.14631	0.24	0.12
…	…	…	…	…	…	…
9	144.72	584	-0.25248	2.82538	12.41	6.12
…	…	…	…	…	…	…
17	609.7	186	0.05408	-0.15302	0.05	0.02
…	…	…	…	…	…	…
38	172.96	497	-0.23386	2.17432	7.50	3.70
…	…	…	…	…	…	…
51	15.36	462	-0.33777	1.91240	6.29	3.10
…	…	…	…	…	…	…
101	1285.7	548	0.49978	2.55597	7.61	3.75
102	18137.48	946	11.6103	5.53437	135.5	66.83
…	…	…	…	…	…	…
138	698.54	308	0.11266	0.75995	0.70	0.35
139	752.64	217	0.14833	0.07896	0.02	0.01
140	809.25	40	0.18565	-1.24560	2.58	1.27
141	178.1	253	-0.23047	0.34837	0.37	0.18
142	81.48	405	-0.29418	1.48585	3.89	1.92
143	1093.86	241	0.37330	0.25856	0.14	0.07
144	1002.76	156	0.31323	-0.37753	0.52	0.26
145	551.88	92	0.01596	-0.85646	1.05	0.52

Table II contains the normalized data from 145 projects, MSD and TS for MSD for normalized data. These data is normalized by S_U family of the transformation (6). In these case the parameters are such: $\gamma_1 = -1{,}448408$, $\gamma_2 = -0{,}489606$, $\eta_1 = 0{,}717501$, $\eta_2 = 0{,}655549$, $\varphi_1 = 71{,}11167$, $\varphi_2 = 1178{,}5237$, $\lambda_1 = 46{,}09214$ and $\lambda_2 = 513{,}9309$. The sample correlation matrix (3) of the **Z** is used as the approximate moment-matching estimator of correlation matrix Σ

$$S_N = \begin{pmatrix} 0.993109 & 0.716010 \\ 0.716010 & 0.993119 \end{pmatrix}.$$

In Table II the last column reveals that projects 4, 17, 101, 102, 138, 140 and 144 are bivariate outliers, since $F_{2,143;0.05}$=3.06. We note, only for two projects 101 and 102 the results are the same in both cases. For other projects, the results of bivariate outliers do not match. First of all, this is due to poor normalization (or normality) of standardized data by formula (7). It is known that Mardia's multivariate kurtosis [9] β_2 equals 8 under bivariate normality. The values of β_2 equal respectively 131.20 and 8.21 for the data from Table I and Table II. These values indicate that the necessary condition for bivariate normality is practically performed for the normalized data from Table II and does not hold for standardized data from Table I by the formula (7).

The prediction ellipses (Fig. 1 and Fig. 2) indicate on the same results. On Fig. 1 and Fig. 2 the standardized and normalized data set for 145 projects and the prediction ellipses are presented. On Fig. 2 the prediction ellipse (5) also reveals that seven data points (projects 4, 17, 101, 102, 138, 140 and 144) are bivariate outliers as in Table II.

TABLE IV. TS FOR MSD FOR THE NORMALIZED DATA

Project	Normalized size	Normalized actual effort	d_i^2	TS for MSD
1	-1.002326	-1.216116	1.52	0.75
2	-1.662999	-0.724990	3.26	1.61
3	1.212616	2.054895	4.40	2.17
4	-1.827636	-0.023010	6.88	3.39
…	…	…	…	…
9	-0.553410	-0.496653	0.33	0.16
…	…	…	…	…
17	0.814075	-1.140640	6.93	3.42
…	…	…	…	…
38	-0.348010	0.648245	1.82	0.90
…	…	…	…	…
51	-2.181756	-1.010318	5.46	2.69
…	…	…	…	…
101	-0.417388	0.556097	9.29	4.57
102	1.350987	2.154155	13.70	6.76
…	…	…	…	…
138	0.923278	-1.354447	9.42	4.64
139	0.982474	-0.178135	2.62	1.69
140	1.039605	-1.331836	10.17	5.02
141	-0.315695	-1.129770	1.81	0.89
142	-1.288339	-1.344194	2.03	1.00
143	1.273260	0.666751	1.76	0.87
144	1.206397	-0.664230	6.40	3.16
145	0.732924	-0.843953	4.49	2.21

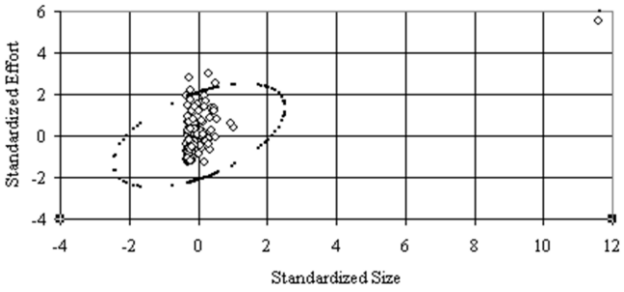


Figure 1. Standardized data set for 145 projects

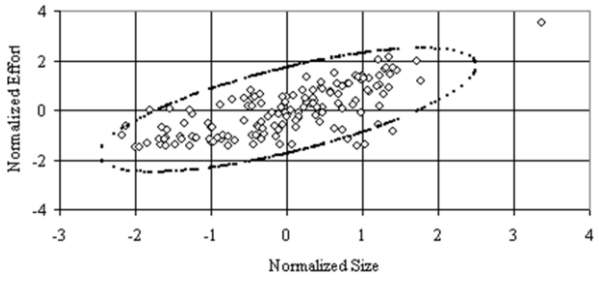


Figure 2. Normalized data set for 145 projects

On Fig. 3 the data set for 145 projects and the transformed prediction ellipse are presented.

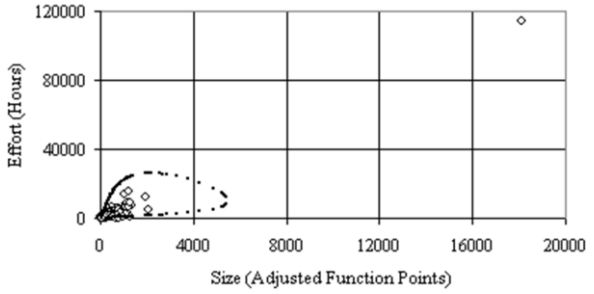


Figure 3. Data set for 145 projects

On Fig. 3 the transformed prediction ellipse also reveals that seven data points (projects 4, 17, 101, 102, 138, 140 and 144) are bivariate outliers. We note, if the anomaly detection technique [10] based on the Grubb test applies for detecting outliers in the normalized data for 145 projects then 144 data sample units do not appear to be an outlier in each of the univariate distributions.

CONCLUSIONS

From the examples we conclude that the proposed technique is promising. For other bivariate non-Gaussian data set of effort and mass of designed the section of the ship from 188 designs of sections the results are similar.

REFERENCES:

[1] D.M. Hawkins, Identification of Outliers, London; New York: Chapman and Hall, 1980, 188 p.

[2] R.A. Johnson and D.W. Wichern, Applied Multivariate Statistical Analysis. Pearson Prentice Hall, 2007, 800 p.

[3] S. Prykhodko, N. Prykhodko, L. Makarova and K. Pugachenko, "Detecting outliers in multivariate non-Gaussian data on the basis of normalizing transformations", unpublished.

[4] S. Prykhodko, N. Prykhodko, L. Makarova and K. Pugachenko, "Multivariate outlier detection technique based on normalizing transformations for non-Gaussian data", unpublished.

[5] A.A. Afifi and S.P. Azen, Statistical analysis: a computer oriented approach, New York; London: Academic Press, 1972, 366 p.

[6] V. Chew "Confidence, prediction and tolerance regions for the multivariate normal distribution" *Journal of the American Statistical Association*, Vol. 61, Issue 315, pp.605-617, 1966.

[7] P.M. Stanfield, J.R. Wilson, G.A. Mirka, N.F. Glasscock, J.P. Psihogios, J.R. Davis "Multivariate input modeling with Johnson distributions", in *Proceedings of the 28th Winter simulation conference WSC'96*, December 8-11, 1996, Coronado, CA, USA, ed. S.Andradyttir, K.J.Healy, D.H.Withers, and B.L.Nelson, IEEE Computer Society Washington, DC, USA, 1996, pp. 1457-1464.

[8] B. Kitchenham, S.L. Pfleeger, B. McColl, and S. Eagan, "An empirical study of maintenance and development estimation accuracy", *The Journal of Systems and Software*, 64, pp.57-77, 2002.

[9] K.V. Mardia, "Measures of multivariate skewness and kurtosis with applications", *Biometrika*, 57, pp. 519–530, 1970.

[10] S.B. Prykhodko, "Statistical anomaly detection techniques based on normalizing transformations for non-Gaussian data", in *"Computational Intelligence (Results, Problems and Perspectives)", Proceedings of the International Conference*, Kyiv-Cherkasy, Ukraine, May 12-15, 2015, pp. 286-287.

Discriminant face features extraction , analysis & its application in multipose face recognization: a survey

Mr.Shailesh S.Shekapure M.E(Scholar), Prof.Niketa V.Kadam, Asstt.Professor
Prof.Ram Meghe Institute of Technology & Research, Badnera, Amravati, India

***Abstract* – As one of the excellent learning and classification performance, SVM and ISVM has become a research topic in the field of machine learning and has been applied in many areas, such as face detection and recognition, handwriting automatic identification and automatic text categorization. Face recognition is a challenging computer vision problem. Given a face database, goal of face recognition is to compare the input image class with all the classes and then declare a decision that identifies to whom the input image class belongs to or if it doesn't belong to the database at all. In this survey, we study face recognition as a pattern classification problem.In this paper, we study the concept of SVM and sophisticated classification techniques for face recognition using the SVM and ISVM along with the advantages and disadvantages. This paper not only provides an up-to-date critical survey of machine learning techniques but also performance analysis of various SVM and ISVM techniques for face recognition are compared.**

Keywords* – *Face Recognition, Machine Learning, Support Vector Machine, Classification,Genetic algorithm

INTRODUCTION

Face recognition is a major issue in the field of pattern recognition, its research contributes to not only the realization of intelligent machines, but also the promotion of the human visual system itself. For a long time, face recognition has gotten earnest concern from the researchers in pattern recognition, artificial intelligence, computer vision, physiology, and other fields, a variety of identification algorithms have been proposed, many commercial face recognition systems have been applied in real world widely[10].There are three main methods of face recognition: structural matching method based on the characteristics, whole matching method and combination method. Geometric characteristics of the face, such as the location ,size, relations of eyes, nose,chin and so on, are used to represent the face in structural matching method; in whole matching method, the gray image of whole face acts as input to train and test the classifier, such as the wavelet-based. Elastic Matching, the principal component analysis and so on; combination method is a combination of the two former methods, usually the overall characteristics is used for a preliminary identification, and then local Features for further identification[10]. However, a major challenge of face recognition is that the captured face image often lies in a high-dimensional feature space. These high-dimensional spaces are too large to allow effective and efficiency face recognition. Due to the consideration of the curse of dimensionality, it is often essential to conduct dimensionality reduction to acquire an efficient and discriminative representation before formally conducting classification. Once the high-dimensional face image is mapped into lower-dimensional space, conventional classification algorithms can then be applied [11].

1 BASIC CONCEPT OF SVM

Support vector machines (SVMs) are a general algorithm based on guaranteed risk bounds of statistical learning theory. A support vector machine (SVM) is a type of state-of-the-art Patten recognition technique whose foundations stem from statistical learning theory. We have found numerous applications, such as in face recognition, character recognition, face detection and so on [6]. In a support vector machine, the direct decision function that maximizes the generalization ability is determined for a two-class problem. Assuming that the training data of different classes do not overlap, the decision function is determined so that the distance from the training data is maximized. We call this the optimal decision function. Because it is difficult to determine a nonlinear decision function, the original input space is mapped into a high dimensional space called feature space. And in the feature space, the optimal decision function, namely, the optimal hyperplane is determined [9]. The basic principle of SVM is to find an optimal separating hyperplane so as to separate two classed of patterns with maximal margin. It tries to find the optimal hyperplane making expected errors minimized to the unknown test data, while the location of the separating hyperplane is specified via only data that lie close to the decision boundary between the two classes, which are support vectors.A common problem that can be observed in many AI engineering applications is pattern recognition. The problem is as follows – given a "training set" of vectors, each belonging to some known category, the machine must learn, based on the information implicitly contained in this set, how to classify vectors of unknown type into one of the specified categories. Support vector machines (SVMs) provide one means of tackling this problem[7].

1.2 Basic Concept of ISVM

Incremental learning is proposed by [Syed, 1999] to solve two typemachine learning problems: one is that computer's memory is not enough or training time too long when training data set is too large; another is we can

obtain the maturity data set at beginning and have to use online learning, that may improve learning precision in the using process with increasing of samples. The key of incremental learning is which learning information should be retrained in the previous training and how to deal with newly adding data set. Syed proposed the incremental SVM learning algorithm at first. The algorithm, first, train training data set and obtain classifier and all support vectors; then we obtain new training data set through merging support vectors and new adding data set; finally, we train new training data set [14]. An approach to support incremental learning is to train the classifier using batches of data subsets, that is to say, only one subset of the data is to be trained at any one time and results subsequently combined. This method is called batch model learning or batch learning, using this method, the learning results are "incremental" combined and deposited. The batch learning methods utilize the property of SVM that only a small fraction of training data end up as support vectors, the SVM is able to summarize the data space in a very concise manner, and assume that the batches of data will be appropriate samples of the data.Clearly, the problem is the learning results are subject to numbers of batches and state of data distribution but always the distribution of data is unknown. In the beginning of the learning process, training dataset is not fully available as in batch learning, data can arrive at any time, so incremental learning algorithms differs from batch ones greatly.

It is proved that the location of the optimal hyperplane is only related with linear combination of support vectors[15].

2 CLASSIFICATION TECHNIQUES

2.1 Face Recognition By SVM's Classification Of 2D And 3D Radial Geodesics

Support Vector Machines (SVMs) are used to perform face recognition using 2D- and 3D-RGDs. Due to the high dimensionality of face representations based on RGDs, embedding into lower-dimensional spaces is applied before SVMs classification[1].Radial geodesicis defined as the particular geodesic that connects one point of themodel to the nose tip along the radial direction connecting the twosurface points.In this technique, an original framework is used to represent 2D and 3D facial data using radial geodesic distances (RGDs) computed with respect to a reference point of the face (i.e., the nose tip). The objective is to define a face representation that can be extracted from 2D face images as well as from 3D face models and used to directly compare them. In 3D, the RGD of a point on the face surface is computed as the length of the geodesic path that connects the point to the nose tip along a radial direction.[1] In 2D, the RGD from a pixel to the fiducial point is computed based on the differences of the image gray level intensities along a radial path on the image.2D radial geodesic distances (2D-RGDs) are computed according to the intensity variations and proximity of image pixels. Matching between 2D- and 3D-RGDs results into feature vectors which are classified by a set of Support Vector Machines (SVMs). Since the feature vectors lay in a high-dimensional space, dimensionality reduction methods are applied before SVMs classification[1].

Advantage: The objective is to define a face representation that can be extracted from 2D face images as well as from 3D face models and used to directly compare them [1]. Three-dimensional (3D) facial data has been exploited as a means to improve the effectiveness of face recognition systems.

Disadvantage: However, a common drawback of solutions that perform recognition by matching 3D facial data is that, despite recent advances in 3D acquisition technologies and devices, acquisition of 3D facial data of a person can be accomplished only in controlled environments and requires the person to stay still in front of a 3D scanning device for a time that ranges from some seconds up to a few minutes [1].

2.2 Facial Expression Classification from Gabor features using SVM :

In this technique facial expressions are analyzed using Gabor features. To reduce the computational complexity, Gabor features are selected in a different manner. For each fixed scale and orientation, a set of Gabor faces are obtained. The Gabor features extracted from different blocks of Gabor faces are used for further analysis. Support Vector Machine is used to classify different expressions [18].Features based on Gabor filters have been used in image processing due to their powerful properties. The main characteristics of wavelet are the possibility to provide a multi resolution analysis of the image in the form of coefficient matrices. These are used to extract facial appearance changes as a set of multiscale and multi orientation coefficients. Gabor filter is shown to be robust against noise and changes in illumination. Gabor kernels are characterized as localized, orientation selective, and frequency selective. The Gabor wavelet representation of images allows description of spatial frequency structure in the image while preserving information about spatial relations [18].Images are divided in to 5 blocks of 28 x 28 sizes. Mean and standard deviation are computed for each sub block.This is considered as feature vectors to an SVM classifier, which is used to discriminate different types of expressions. Initially one expression group is selected. All the images under this group are classified as +1 and others as -1.This iteration process continues until all the expression groups are classified properly [18].

Advantage: The main characteristics of wavelet are the possibility to provide a multi resolution analysis of

the image in the form of coefficient matrices [18]. These are used to extract facial appearance changes as a set of multi scale and multi orientation coefficients. The Gabor wavelet representation of images allows description of spatial frequency structure in the image while preserving information about spatial relations. This method improves both the processing speed and efficiency[18].

Disadvantage:Evaluating filters to convolve the face image is quite time consuming.

3 ISVM FOR FACE RECOGNITION

When the dataset is too large to be fit in the memory at once or to be read into the memory, SVM suffers problem to train the dataset. In such a case online learning should be used if the data set is not at all obtainable in the beginning. So incremental learning is proposed to solve such machine learning problems like large memory requirement and large training time when the dataset is very large. At first the algorithm trains the current dataset and construct the initiatory optimal hyperplane i.e. read the entire support vector[22]. Then the algorithm checks the current dataset for the positive and negative group. If the current dataset lies in the positive group then the input feature vectors are matched with the output, if correct matching is found then success n return otherwise increment the value of lagrangian parameter up to 10% and continue this process until correct classification is achieved. If the current dataset lies in the negative group then again the input feature vectors are matched with the output, if correct matching is found then success n return otherwise the value of lagrangian parameter is decremented up to 10% and continue this process until correct classification is achieved[22].

Advantage*:* Recognition rate achieved 100%.

Disadvantage *:* Matching time is little bit larger than SVM.

CONCLUSION

We have present a survey of face recognition based on SVM. This extensive survey has addressed how SVM can be applied for recognition of faces and how the performance of recognition system can be improved using SVM. The performance, concepts along with advantage and disadvantage of SVM techniques for face recognition are summarized in this paper.

REFERENCES

[1] Stefano Berretti, Alberto Del Bimbo, Pietro Pala, Francisco Jos`e SilvaMata ,"Face Recognition By SVM Classification Of 2D And 3D Radial Geodesics" University of Firenze, Italy IEEE 2008.

[2] Nils J. Nilsson, "Introduction to Machine Learning", Department of Computer Science Stanford University, December 1996.

[3] Jun-Ying Gan, Si-Bin He, "Face Recognition Based on 2DLDA and Support Vector Machine", Proceedings of the 2009 International Conference on Wavelet Analysis and Pattern Recognition, Baoding, IEEE 2009, pp.211-214.

[4] Praseeda Lekshmi V, Dr. M Sasikumar, Divya S Vidyadharan "Facial Expression Classification from Gabor features using SVM".

[5] Du Hongle, Teng Shaohua, Zhu Qingfang, "Fast SVM IncrementalLearning Based on Clustering Algorithm", College of Mathematics Luoyang normal university Luoyang, China, IEEE 2009, pp.13-17.

[6] Qian-Ying Chen, Qiang Yang' "Segmentation Of Images Using Support Vector Machines" Proceedings of the Third International Conference on Machine Learning and Cybernetics, Shanghai, 26-29 August 2004

[7] M. Palaniswami, A. Shilton, D. Ralph and B.D. Owen " Machine Learning Using Support Vector Machines " The University of Melbourne, Victoria-3101, Australia.

[8] Ethem Alpaydin, "Introduction To Machine Learning", The MIT PressCambridge, Massachusetts London, England, 2004 MassachusettsInstitute of Technology.

[9] Shigeo Abe, "Support Vector Machines for pattern Classification" , Springer-Verlag London Limited 2005

[10] Kebin Cui, Yingshuag Du, "Application of Boolean Kernel Function SVM in Face Recognition",2009 International Conference on Networks Security, Wireless Communications and Trusted Computing

[11] Xia Sun, Qingzhou Zhang, Ziqiang Wang, " Face Recognition Based onNMF and SVM ", 2009 Second International Symposium on ElectronicCommerce and Security.

[12] Yi - Min Wen and Bao - Liang Lu, "Incremental Learning of Support Vector Machines by Classifier Combining", Department of Computer Science and Engineering, Shanghai Jiao Tong University

[13] Stefan R.Uping. "Incremental Learning with Support Vector Machines ", Department of Computer Science, AI Unit University of Dortmund, Germany. ICDM 2001

[14] Du Hongle, Teng Shaohua, Zhu Qingfang, "Fast SVM Incremental Learning Based on Clustering Algorithm", IEEE2009, pp. 13-17.

[15] Yuanzhi Wang, Fei Zhang, Liwei Chen, "An Approach to Incremental SVM Learning Algorithm", 2008 ISECS International Colloquium on Computing, Communication, Control, and 2008 ISECS International Colloquium on Computing, Communication, Control, and Management, IEEE 2008, pp.352-354.

[16] Aouatif Amine, Sanaa Ghouzali, Mohammed Rziza, Driss Aboutajdine,"Investigation of Feature Dimension Reduction based DCT/SVM for Face Recognition", 2008 IEEE, pp. 188-193.

[17] Guo-Yun Zhang, Shi-Yu Peng, Hong-Min Li "Combination Of Dual- Tree Complex Wavelet And SVM For Face Recognition", Proceedings of the Seventh International Conference on Machine Learning and Cybernetics, Kunming, 12-15 July 2008

[18] [Praseeda Lekshmi V, Dr. M Sasikumar, Divya S Vidyadharan "Facial Expression Classification from Gabor features using SVM".

[19] Zhong-We Li, Jian-Pe Zhang, Jing Yang, "A Heuristic Algorithm To Incremental Support Vector Machine Learning", Proceedings of Third International Conference on Machine Learning and Cybernetics, Shanghai, IEEE 2004, pp.1764-1767.

[20] Jin-Long An, Zhengou Wang, Zhen-Ping Ma, "AnIncremental Learning Algorithm For Support Vector Machine", Proceedings of the Second International Conference on Machine Learning and Cybernetics, San, IEEE 2003, pp.1153-1156.

[21] Pei Jiang, Yongjie Li, "A Hybrid Face Recognition Algorithm Based on WT, NMFs and SVM", University of Electronics Science and Technology of China, IEEE 2008, pp.734-737.

[22] Deepti Sisodia, Shailendra Kumar Shrivastava, R.C. Jain, "ISVM for Face Recognition" ,2010 International Conference on Computational Intelligence and Communication Networks, 2010 IEEE, pp.554-559

Analysis of the data collection mobile user to create technology generations recommendation reports

Andrey Skrypnyk[1], Helen Khalus [2]
[1]undergraduate, NTUU “KPI”. Skrypnik.andrew@gmail.com
[2] senior teacher, NTUU "KPI" http://www.asu.ntu-kpi.kiev.ua/Kafedra/teachers/Halus_O_A/

Lifelogging (Eng. Life-logging, and log on life - "life" and "logbook") - automatic recording of daily life in the digital media using compact portable technical devices and systems. The problems concerned lifelogging many modern companies and corporate projects. But all they solve only some problems and focused on specific corporate requirements. Custom solutions focused on static parameters and view in graphs.

But technology and a very promising direction lifelogging - allows more faithful forecasts for physicians and insurance companies. For the employer - indicators of life balance between work and rest. Purpose and objectives of the study is research is to develop a recommender system based on user context information. This article discusses the first stage - the collection and analysis of the static data.

Keywords: expert system; processing contextual user data; algorithm to generate recommendations; the collection of personal data of the user; iOS application; web server

1. Introduction

The problem is large noisiness data and the small number of public tools for its complex processing tasks within user behavior analysis.

Existing applications and systems offer only a static analysis of the data in graphs and all recommendations based on user count metrics and compare them with expert estimates.

Software technology designed to solve the problem analysis and recommendations as to individuals and companies that analyze the behavior of their employees during the working day, but you first need to understand the quality of incoming data to analyze and assess the correctness of the data.

2. The problem

Modern mobile devices are sufficiently accurate sensors for counting steps, movement, location. They have already partially processing the data of mobile applications, but the algorithms they use - trade secrets and know how we get the correct data, known as measurement error and other statistical data that would help us to assess the correctness of the data.

The data will be transmitted via wireless networks Edge, 3G and WiFi [1].

Additionally, please note that the phone's sensors can greatly affect the amount of energy used to operate the device. Therefore, we must implement three modes of application - Low, Middle, Accuracy.

Low - data is collected only when user location significant changed.

Middle - user data is collected every 120 seconds.

Accuracy - user data is collected every 10 seconds during walking and cycling every 30 seconds during car trips every 120 seconds in a stationary position.

3. The algorithm for data collection of mobile Transporting the server, Conflict Reconciliation and statistical

STEP 1. The user chooses one of three modes of application (Low, Middle, Accuracy) for saving phone during the process of gathering information.

Step 2. Collect data to a mobile device from a variety of sources to which you have access.

STEP 3. Stack collects data before sending to optimize server load.

STEP 4: Making the initial processing, remove the conflict, to reduce noise.

STEP 5. Normalize the absolute value data.

STEP 6. When the stack limit is achieved - send data processing to the server.

Step 7. When receiving data - server decides further merge-conflict challenges and stores data.

Step 8. When user request data - server generates user activity, distributing source data types and categories (sleep, work, activities, entertainment and travel).

STEP 9. Server begins background analytics to process user data to find patterns between his behavior and some basic patterns of behavior.

Step 10. After finding differences for a particular user - server generates recommendations for leveling differences between current behavior and the selected pattern.

STEP 11. Mobile application, if the server has recommendations - displays them to the user.

Step 12. The result of the algorithm is structured data in JSON format.

4. Results

Thus, said collection and processing algorithm was implemented in the iOS [2] mobile application and .NET application server. The algorithm was tested on iPhone 7 in different environments and has the following results. In testing attended by 4,000 people. Residents of Kyiv. Men (70%) and women (30%) aged 18-24 - 12% of 25-34 - 74%, 35-54 - 13%. More than 65% are interested in business and technology.

4.1. Career

Figure 1 shows the total amount of time people spend at work on weekdays, less time people spend on lunch.

Average time: 5:40

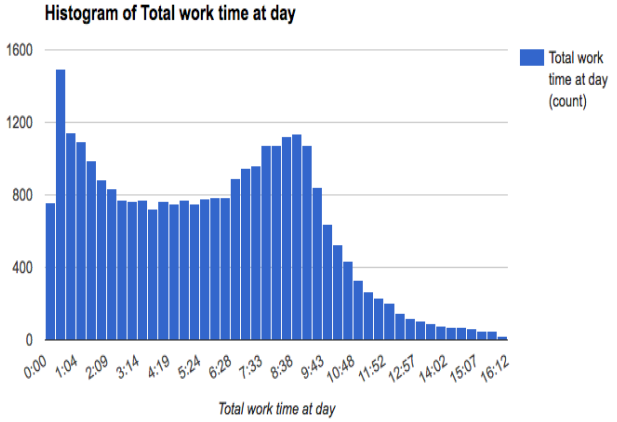


Fig. 1. The distribution of working hours during the day

Figure 2 shows the amount of time people spend at work on weekdays, without interruption.

Average time: 3:12.

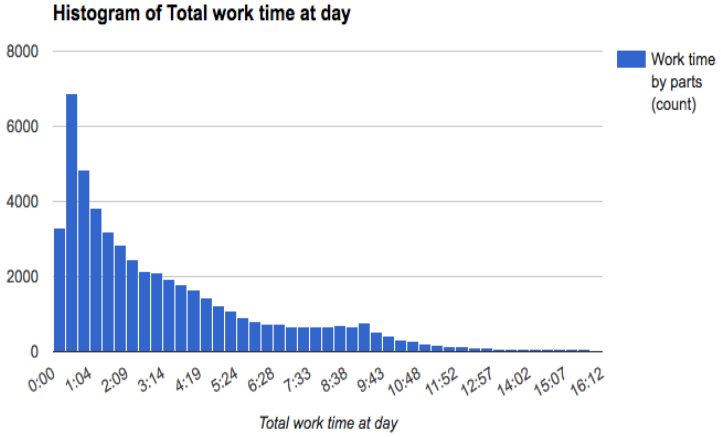


Fig. 2. The distribution of working time during the day in parts

Figure 3 shows the hour when people come to work.

Average time: 10:59.

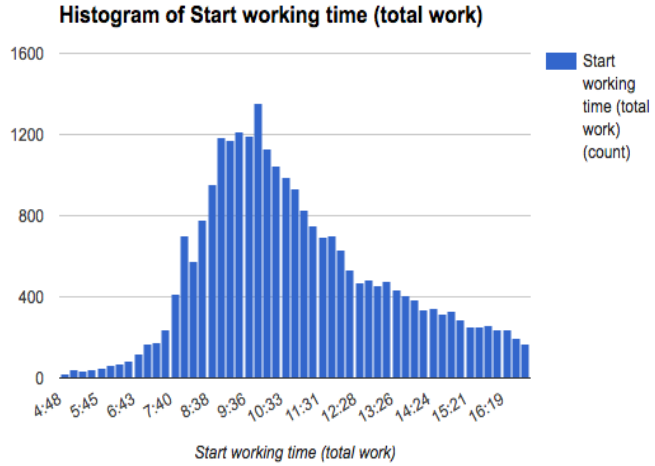


Fig. 3. Start time working hours during the day

Figure 4 shows the hours when users are leaving work.

Average time: 18:17

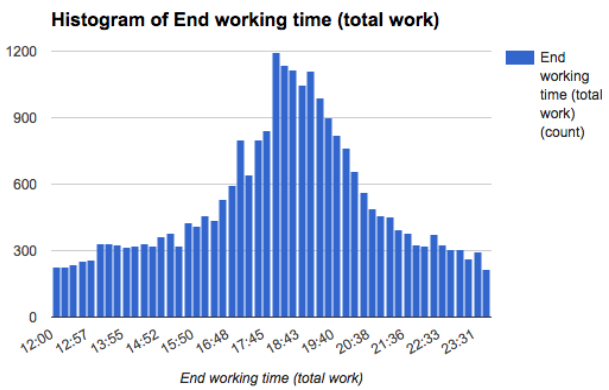


Fig. 4. End time working hours during the day

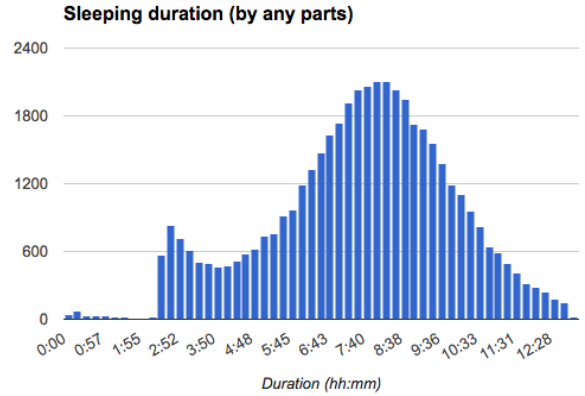


Fig. 5. Average length of time sleeping during the night

4.2. Sleep

Figure 5 shows the time that users are asleep.
Average time: 7:37

4.3. Active

Figure 6 shows the average time of activity a day.

Average time: 1:53

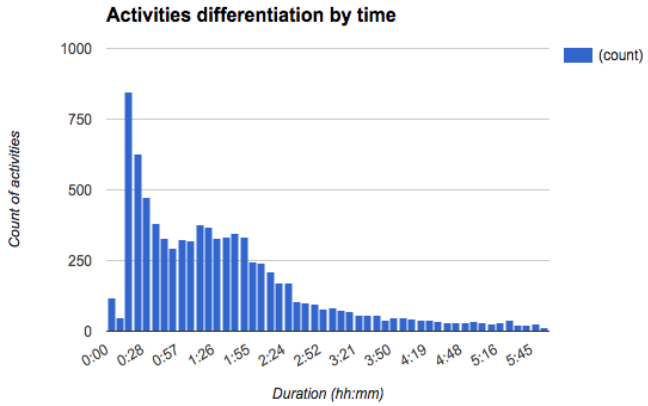


Fig. 6. Average duration of sleep some time intervals

Figure 7 shows a high amount of activity every day.

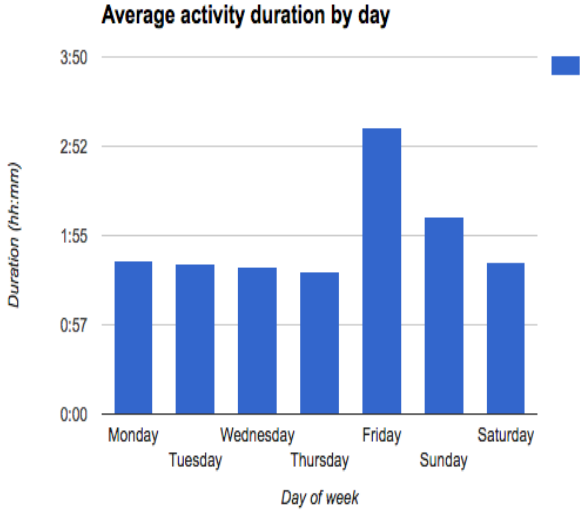


Fig. 7. Distribution of mean time activity by day

5. Conclusions

Thus, the algorithm is implemented in the iOS mobile application and server for .NET [3]. The graph shows quite a lot of data, the truth of which you want to check. Records the amount of time of less than 4 hours per day, sleep duration more than 10 hours before the end of the 14:00 pm.

The future plans - to continue to develop and improve the algorithm, in particular to develop an interactive system confirming the correctness of the data by the user. It is also planned to reduce the% error in data collection devices.

References

[1] Sustainable Energy-Efficient Wireless Applications Using Light Mohsen Kavehrad, Pennsylvania State University, IEEE Communications Magazine, December 2010.

[2] Apple iOS Documentation. [Electronic resourse] – Electronic data. – Apple.com April 2017 – Mode of access: World Wide Web: https://developer.apple.com/library/content/documentation/UserExperience/Conceptual/LocationAwarenessPG/CoreLocation/CoreLocation.html (viewed on April 5, 2017). - Apple, Getting the User's Location

[3] Microsoft Developer Documentation. [Electronic resourse] – Electronic data. – Microsoft April 2017 – Mode of access: World Wide Web: https://docs.microsoft.com/en-us/dotnet/ (viewed on April 10, 2017). - .NET Documentation | Microsoft Docs.

SESSION 6
COMPUTER NETWORKING AND TELECOMMUNICATIONS

Computing Resources Scaling Survey

Y. Kulakov, R. Rader
National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", roman.rader@gmail.com

***Abstract* – The results of the survey about usage of scalable environment, peak workloads management and automatic scaling configuration among IT companies are presented and discussed in this paper. The hypothesis that most companies use automatic scaling based on static thresholds is checked. The insight into the most popular setups of manual and automatic scalable systems on the market is given.**

***Keywords* – cloud computing, auto scaling, computer cluster, survey.**

I. INTRODUCTION

In the modern world, the accelerating business demands on websites make the reliability and efficiency of managing resources crucial. It implies optimization of all the processes in mature companies. Business keeps increasing the dependence on the web applications; workloads and numbers of servers in the companies are growing. In these conditions, any optimization can influence the general system efficiency. In this paper, the optimization of the computing resources is considered. Obviously, the optimization should reduce costs without significant quality loss. In the conditions of unstable and unpredictable peak workloads, cloud computing can significantly reduce costs, making possible to scale up and release unused resources rapidly [1].

By the definition given by NIST, cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction [2]. By this definition, one of the key features of cloud is the scalability, which is the ability to expand and add resources dynamically, to be scaled up to meet demand through replication and distribution of requests across a pool or farm of servers [3].

From the application perspective, although cloud allows to flexibly managing resources, the control and monitoring units should be present in the system.

Auto scaling or dynamic scaling implies the automatic control of the action of provisioning or de-provisioning web server virtual machine instances, a dynamic scaling algorithm based on relevant threshold or scaling indicator of the web application is developed [3].

Scaling indicators mentioned above and the approach to the control unit can vary. While the specific method usage depends on the application being developed, set of most popular approaches is described in [4] and covers most of systems (see section III).

In order to understand the current situation on the production systems serving websites and cloud services, this study has been conducted. Main questions to be answered in this study are: (1) how many companies are using scalable environment, (2) are they using its potential (automatic scaling) if they do, (3) how usually companies manage with unexpected peak loads, (4) how do they configure the auto scaling algorithm if applicable. The hypothesis to be check is: (1) most companies deploying in the scalable environment are using automatic scaling; (2) the resource control system is based on static thresholds; (3) thresholds are set empirically without a research.

II. METHODOLOGY

The survey was conducted online using Google Forms [5] service and distributed among software developers and system administrators community in Ukraine. Respondents were invited to answer the questions in chat rooms dedicated to DevOps and system administration topic.

Respondents were asked about the configuration of their clusters; essentially three main issues were to be uncovered: (1) ownership of computing resources; (2) resources behavior in edge cases; (3) the way of configuration of this behavior.

Questions were ordered from broad to narrow in order to find all the details of as many as possible setups.

A. Ownership of computing resources

Basically, to have an ability to optimize the cluster workload during off-peaks, a scalable environment is required. It can be any kind of cloud: the most affordable for any company is public cloud providers help companies to share resources with other companies. Most of the popular public cloud providers bill only the resources you use, so it's possible to save money from idle resources in off-peak hours by disposing of resources back to the pool.

Hence, to understand the share of companies using scalable environment it's important to know the way companies own their resources. In the survey respondents were asked whether they rent the computing resources or own them. The "rent" was divided into two options: (1) rent the physical machines and (2) rent the virtual resources from cloud providers.

B. Resources behavior in edge cases

Under the "edge case" the peak load condition in the system is meant here. Computer systems in the production environment should be configured to handle

different kinds of loads: peak and off-peak hours, increasing and decreasing of load during holidays, even stand the DoS attacks. To understand how companies manage with expected and unexpected peak workloads, they were asked to provide the information about their systems configuration, whether they provision resources manually or automatically, what approach is used.

C. The way of configuration of this behavior

As mentioned above, the approaches of managing the resources can vary. Manual control of the resources implies the presence of the monitoring system, which is used to make a decision about provisioning additional resources or releasing them by a human.

If the automatic resources control approach is used, the control system should be configured to scale the specific application on the specific environment.

Respondents were asked in what way their control unit was configured, whether some research was done to get the parameters for the system, parameters were set empirically by the system administrator, or their control unit doesn't require the configuration.

Also, respondents who use manual resource control approach were asked if they have any plans for a transition to the automatic system.

D. Confidence interval calculation

To be able to compare the values and show the statistically significant percentage we will calculate the confidence interval. Given sample size n and sample proportion $\hat{p}$, which is the percent of respondents in our case, the confidence interval for p - the population proportion of the survey responses for a single option of the question.

Question options	x_1	x_2
Number of answers	n_1	n_2

where $n = n_1 + n_2$.

The dispersion of the sample is given by [6]:

$$D_s = \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n} \quad (1)$$

Substituting n=2 and the values, the dispersion of the sample would be:

$$D_s = \frac{n \cdot n_i - n_1^2}{n} \quad (2)$$

The sample proportion $\hat{p}$ is:

$$\hat{p} = \frac{n_1}{n} \quad (3)$$

Substituting (3) into the (2):

$$D_s = \hat{p} \cdot (1 - \hat{p}) \quad (4)$$

The correction of the sample dispersion to calculate the population dispersion is not required since the population is greater than 30, $D_s \approx D_p$ [6] (see the III.A section).

Hence, the confidence interval is given by:

$$\hat{p} \pm t \cdot \sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{n}} \quad (5)$$

For the confidence level of 95%, the value of $t = 1.06$.

III. Results

A. General numbers

Based on the data from State Statistics Service of Ukraine on 2015, there are 13617 companies registered in Ukraine by the "information and telecommunications technology" activity [7].

44 companies working in Ukraine responded to the survey. 41% of them are Ukrainian companies and 59% of them are official representatives of foreign companies in Ukraine.

B. Resources ownership

Among responded companies, 61.4% rent computing resources and 38% use their own hardware. Moreover, 96.3% of companies who rent resources, are renting them from cloud providers and only 3.7% are renting dedicated servers (see fig. 1).

TABLE I. OWNERSHIP OF RESOURCES

	Value and 95% confidence interval		
Ownership	***Own hardware***	***Rent computing resources***	***Total***
Responses in the survey	17 answers	27 answers	44 answers
Among IT companies	38.6%±14%	61.4%±14%	100%
Type of rented resources (among renting group)	***Cloud providers***	***Dedicated servers***	***Total***
Responses in the survey	26 answers	1 answer	27 answers
Among IT companies	96%±7%	4%±7%	100%

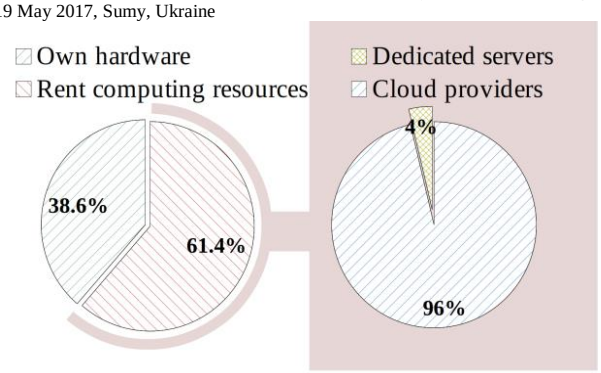


Figure 4. Ownership of resources

C. Behavior on edge cases

In this section, the results of the survey about the computer systems configuration are considered.

TABLE II. RESOURCES MANAGEMENT ON EDGE CASES

	Value and 95% confidence interval		
	Automatic	***Manual***	***Total***
Responses in the survey	18 answers	26 answers	44 answers
Among IT companies	40.9%±14%	59.1%±14%	100%

To get the insight in these two groups the approaches of resource management in each group were considered (see table III and table IV).

Respondents were asked how they act in conditions when their system experience higher workload than expected. There were two majorities of answers (see table II), (1) any kind of automatic resource management and (2) manual. Also, two respondents answered that their workload is highly predictable. For this survey purposes, these answers were assigned to the group of manual resource management.

TABLE III. TYPE OF MANUAL RESOURCES MANAGEMENT

	Value and 95% confidence interval			
	Manual management	***Queues and metrics thresholds as indicators***	***Regression model for workload prediction***	***Total***
Responses in the survey	17 answers	7 answers	2 answer	26 answers
Among IT companies	65.4%±18%	26.9%±17%	7.7%±10%	100%

TABLE I. TYPE OF AUTOMATIC RESOURCES MANAGEMENT

	Value and 95% confidence interval			
	Static thresholds	***Serverless architecture***	***Queueing theory***	***Total***
Responses in the survey	14 answers	3 answers	1 answer	18 answers
Among IT companies	77.8%±19%	16.7%±17%	5.6%±11%	100%

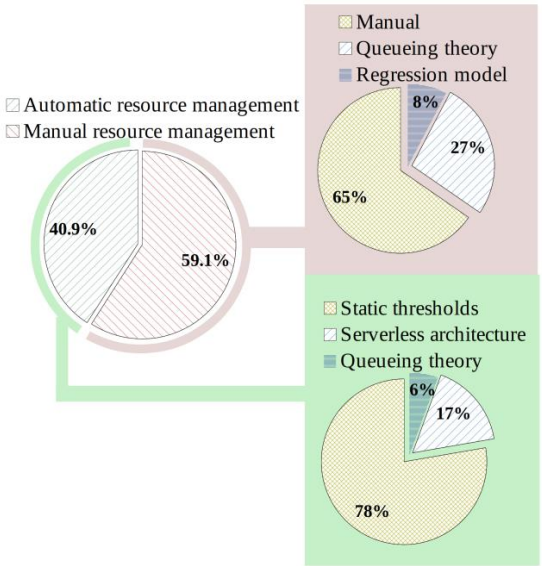


Figure 5. Resources management

D. Automatic resource management configuration

The automatic resource control unit may require a configuration to work in specific environment with concrete application. The ways of setting configuration parameters are considered in table V.

TABLE II. CONFIGURATION APPROACH

	Value and 95% confidence interval				
	Research has been done	***Parameters set empirically***	***Load testing***	***Nonparametric algorithm***	***Total***
Responses in the survey	8 answers	5 answers	1 answer	4 answers	18 answers
Among IT companies	44.4%±22%	27.8%±20%	5.6%±10%	2.2%±7%	100%

E. Trends of future

Respondents were asked whether they are in process of migration to auto-scaling solutions now or not. Among those who are not using auto scaling currently, 39% (±14% among all IT companies) of respondents answered they are in process of development or adoption the auto-scaling control unit for their systems.

IV. DISCUSSION

Main questions this survey was intended to answer are (1) how many companies are using scalable environment, (2) are they using its potential (automatic scaling) if they do, (3) how usually companies manage with unexpected peak loads, (4) how do they configure the auto scaling algorithm if applicable. Also, the hypothesis was put forward to check that (1) most companies deploying in the scalable environment are using automatic scaling; (2) the resource control system is based on static thresholds; (3) thresholds are set empirically without a research.

A. How many companies are using scalable environment?

Considering the results that 61.4% are renting resources and 96% of them are renting them from cloud providers, the answer is 59% of companies have a scalable environment (see table I). This result is consistent with study by RightScale in 2017 [8], considering the confidence interval and focus of this study on the Ukrainian companies.

B. Are companies using the potential of auto-scaling and how they manage with peak loads?

Among respondents, 40.9% are using auto-scaling (see table II). Considering that 59% have a scalable environment (see IV.A), we can state that majority of companies (70%) are using the potential of the auto-scaling environment, which confirms the (1) of the hypothesis. Moreover, most of them (77.8%) are using the method of static thresholds for their resource control systems (see table III), which confirms (2) of the hypothesis.

The amount of respondents is the limitation to get statistically significant results of using less popular auto-scaling techniques than usage of the static thresholds approach.

C. How usually companies manage with unexpected peak loads?

The study showed that less than half of companies set the parameters of their auto-scaling system empirically (see table V). The majority (44.4%) made some kind of research to get them. This result rejects the (3) of the hypothesis. Also, since customers are ready to make a research to configure the auto-scaling system, for researchers that mean that customers need the balanced solution that both easy to setup and has flexible, easily understood configuration parameters to make the auto-scaling system adaptable for a specific application.

CONCLUSIONS

The purpose of the study is described, main questions to answer by the study are put and the hypothesis is articulated. The methodology of conducting the survey is given. The results with calculated margins of errors of the conducted survey are presented in this paper. Results summarize and interpretation of them are described, the limitations of given results are described. The hypothesis is confirmed by two items and rejected by one, interpretation is provided.

REFERENCES:

[1] T. Dillon, C. Wu, and E. Chang, "Cloud computing: issues and challenges." 2010 24th IEEE International Conference on Advanced Information Networking and Applications (AINA), April 2010.

[2] P. Mell, and T. Grance, "The NIST definition of cloud computing." National Institute of Standards and Technology, Special Publication 800-145, 2009.

[3] T. C. Chieu, A. Mohindra, A. A. Karve, and A. Segal, "Dynamic scaling of web applications in a virtualized cloud computing environment", 2009 IEEE International Conference on e-Business Engineering, pp. 281-286, October 2009.

[4] T. Lorido-Botrán, J. Miguel-Alonso, and J. A. Lozano, "Auto-scaling techniques for elastic applications in cloud environments." Department of Computer Architecture and Technology, University of Basque Country, Tech. Rep. EHU-KAT-IK-09-12, 2012.

[5] R. Rader "Google Forms: Computing Cluster Scaling Survey 2017", docs.google.com/forms/d/e/1FAIpQLSfYzDhmRU3_NebtCep77DlZmVjpDFlCkUB2k1uHhjYIcAuhBg/viewform, accessed April 15, 2017.

[6] V. E. Gmurman "Probability theory and mathematical statistics." Moscow: Vysshaya Shkola, 1972 (in Russian).

[7] State Statistics Service of Ukraine "Number of business entities by economic activities in 2015" ukrstat.gov.ua/operativ/operativ2014/fin/osp/ksg/ksg_u/ksg_u_15.htm, 2015, accessed April 15, 2017 (in Ukrainian).

[8] K. Weins, "Cloud Computing Trends: 2017 State of the Cloud Survey", www.rightscale.com/blog/cloud-industry-insights/cloud-computing-trends-2017-state-cloud-survey, February 2017, accessed April 17, 2017.

Verification of Cryptosystems Sustainability as the Main Criterion for Development of Common Information Security Policy

Andrii Boiko[1], Vira Shendryk[1], Lina Cherednichenko[2]
[1]Sumy State University, 2, Rymskogo-Korsakova st., 40007 Sumy ,Ukraine
[2]George Brown College, Toronto, Canada

Security of data resources has become one of the main issues of modern society. Encryption is one of the most reliable ways to protect data from unauthorized disclosure. One of the key factors that influenced the formation of a new approach to information security is a significant growth of distributed-processing systems and use of computer networks for communication between users. It is a key reason to develop an information security policy.

Keywords – symmetric cryptosystem, information security policy, public key cryptosystems, requirements for cryptosystems, Kerckhoffs' principle.

I. INTRODUCTION

The processes that take place today in the world directly have influence on the information security of any company. At the same time there are new factors which need to be considered while verifying the actual state of information security and identifying key issues and trends in this area.

Today, one of the main parts of the overall security is the information security. The evolution dynamics of information technology in social and economic sphere requires a comprehensive approach to addressing information security. To ensure complete, accurate and timely information there is a need to provide security functions in information system as well as to secure the information resources.

Thus, the development of a comprehensive information security policy which throughout the information life cycle determines the relevance and timeliness of this study.

II. STRUCTURE OF CRYPTOSYSTEMS

The analysis of cryptosystems makes it possible to identify the main areas of their application. The main functions are:

- Provision of confidential information for further data transfer through communication channels (e.g. e-mail);
- Authenticity of transmitted messages;
- Data storage (documents, databases) on storage medium in encrypted form.

The cryptographic methods are an integral part of the information security policy. The term cryptographic methods of information protection means the special methods of encryption, encoding or other transformation of information, which makes the data content inaccessible without a key cryptograms and reverse transformation. The cryptographic method is one of the most reliable methods of security, as it secure information itself, and not access to it.

Modern cryptosystem includes four major sections:

- Symmetric cryptosystem;
- Public key cryptosystems;
- Digital signature;
- Key management.

Symmetric cryptosystems are based on such encryption methods where encryption and decryption use the same key (Fig. 1).

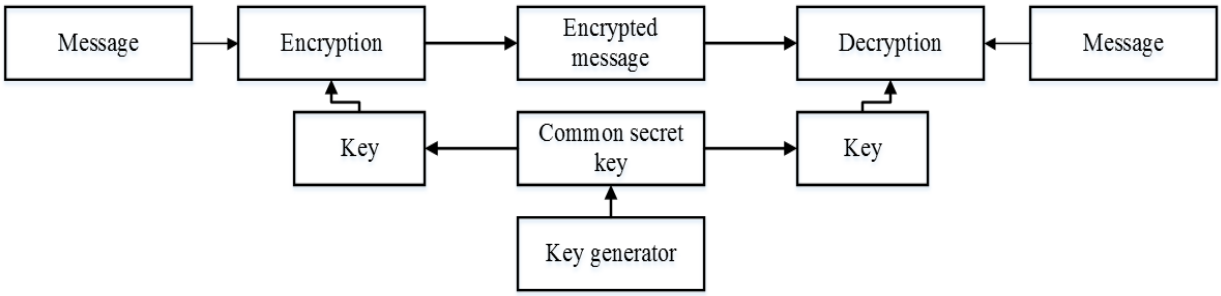


Figure 1. Symmetric cryptosystems

Public key cryptosystems use two keys: one is public key and another one is private key, these two keys are mathematically related to each other. Data is encrypted by using a public key that is available to any person and decrypted by using a private key known only to the recipient (Fig. 2).

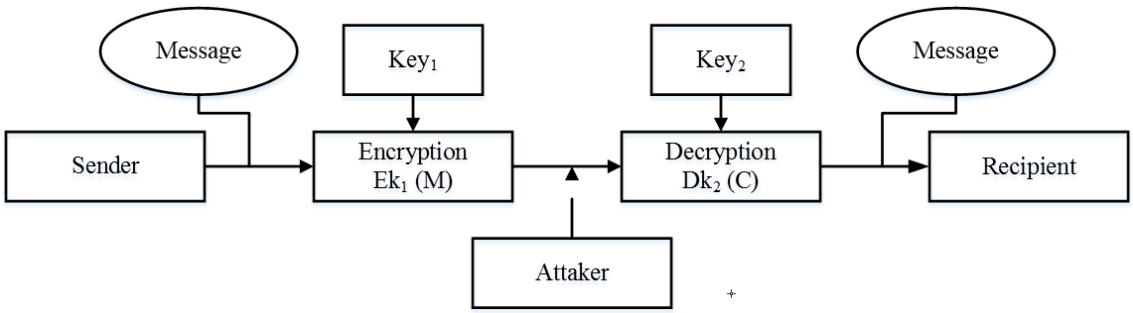


Figure 2. Public key cryptosystems

The system of digital signature is its cryptographic transformation, which are attached to the text and gives a recipient possibility to check the authorship and authenticity of the message.

The process of key management in information processing system consists of formation and distribution.

III. REQUIREMENTS FOR CRYPTOSYSTEMS

The process of cryptographic data closure can be implemented both programmatically and in hardware. Hardware implementation is significantly more expensive, but it also has advantages: high performance, simplicity, security, etc. The software implementation is more practical, allows for a certain flexibility in use. For modern cryptographic information security systems, the following requirements should be emphasized:

- The encrypted message must be readable only if there is a key;
- The number of operations necessary to determine the encryption key used for the fragment of the encrypted message and the corresponding plaintext corresponding to it should be not less than the total number of possible keys;
- The number of operations which is necessary to decrypt information by searching all possible keys must have a strict lower bound and go beyond the capabilities of modern computers;
- Knowledge of the encryption algorithm should not affect to the reliability of protection;
- A minor change in the key should lead to a significant change in the type of encrypted message, even when using the same key;
- The structural elements of the encryption algorithm must be unchanged;
- The length of the encrypted text should be equal to the length of the source text;
- Any key of the set must provide reliable protection of information.

IV. VERIFICATION OF CRYPTOSYSTEMS SUSTAINABILITY

To verify the sustainability of data security cryptosystems it is necessary to follow Kerckhoffs' principle: A cryptosystem should be secure even if everything about the system is public knowledge. Therefore, the analysis of sustainability of cryptosystem is based on the assumption that the opponent knows the detailed description of the system, statistical properties of the message language, the space of possible keys and cryptograms. Also, he may have some information about the context of the message, etc. The only thing the offender mustn't know is the secret cryptographic key used by users of secure cryptographic systems. The system of analysis of cryptographic algorithms can be divided into two subsystems:

- secure cryptographic subsystems with different classes of ciphers;
- subsystems of cryptanalysis.

Typically, different approaches are used to verify the sustainability of secure cryptographic systems; the most interesting among them are the information-theoretical, complexity-theoretic and system-theoretical approaches.

According to information-theoretical approach used to verify the sustainability of cryptographic systems, the cryptosystems can be divided into absolute stable and relatively stable. Sustainability of cryptographic systems with absolute stability does not depend on any abilities of intruder and cannot be decreased under any circumstances. Sustainability of cryptographic systems with relative stability depends on the abilities of opponent and his methods which can vary depending on various factors.

Consider the example of an arbitrary function y = f (x), which is shown graphically (Fig. 3).Suppose that we have a set X = {a, b, c, d, e} and a set Y = {1,2,3,4,5}. Note that, the function is defined by two sets X and Y, and by the rule f, which assigns one element from the set Y to each element of the set X. The set X is called the function domain, and the set Y is the value domain.

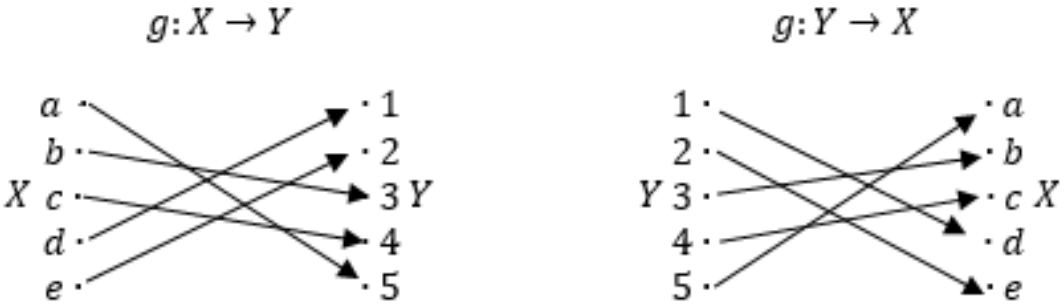


Figure 3. Bijective function f and its inverse function g= f^(-1)

The element y of the set Y is a direct image of the element x, and the element x is an inverse image of y. The mapping of elements from the set X to the set Y is written as follows: f: X → Y.

The set of all elements y having at least one inverse image is called the direct image of the function f and is denoted by Im (f).

A function is said to be single-valued function (one-to-one mapping) if every element of the set Y is the direct image of not more than one element of the set X. A function f is said to be a bijection function if it is single-valued function and Im (f) = Y. A function of the form g = f ^ (-1) is said to be the inverse function of f. Among bijective functions there is a class of functions which is called involutions which are most often used to build secure cryptographic systems.

A bijective function is called an involution if the function domain coincides with the value domain, that is, X = Y = S, and also the inverse function coincides with the function f = f -1. Fig. 3 shows an example of an involution for the set S = {1,2,3,4,5}.

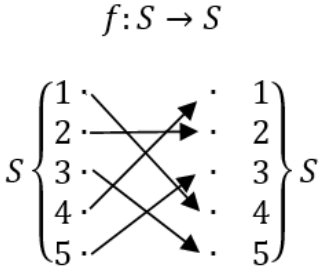


Figure 4. Involution f for the set S={1,2,3,4,5}

The existence of an inverse function is the basis for building of data encryption systems, by means of which cryptograms can be precisely decrypted into messages. Consistent application of encryption function at first, and then decryption function to an undefined message x∈S precisely restores this message: f (f (x)) = x.

V. CONCLUSIONS

The clarification of the question whether the cryptosystem is absolute or relative stable constitutes an important task of the information-theoretical approach used to verify the sustainability of data security cryptosystems. If, within the information-theoretical approach, the cryptosystem is considered as absolute stable, then the degree of its sustainability should be further verified using the complexity-theoretic and system-theoretical approaches. It should be noted that the information-theoretical approach is often referred to the class of theoretical approaches used to verify the sustainability of cryptosystems, and the rest referred to the class of practical approaches.

REFERENCES:

[1] Siponen, M. and Willison, R. (2009). Information security management standards: Problems and solutions. Information & Management, 46(5), pp.267-270.

[2] Kenkre, P., Pai, A. and Colaco, L. (2015). Real Time Intrusion Detection and Prevention System. Advances in Intelligent Systems and Computing, pp.405-411.

[3] Boiko, A. and Shendryk, V. (2017). System Integration and Security of Information Systems. Procedia Computer Science, 104, pp.35-42.

[4] Eskander, G., Sabourin, R. and Granger, E. (2014). Improving Signature-Based Biometric Cryptosystems Using Cascaded Signature Verification-Fuzzy Vault (SV-FV) Approach. 2014 14th International Conference on Frontiers in Handwriting Recognition.

[5] Chadha, R., Cheval, V., Ciobâcă, Ş. and Kremer, S. (2016). Automated Verification of Equivalence Properties of Cryptographic Protocols. ACM Transactions on Computational Logic, 17(4), pp.1-32.

[6] Zhao, T., Ran, Q., Yuan, L., Chi, Y. and Ma, J. (2016). Information verification cryptosystem using one-time keys based on double random phase encoding and public-key cryptography. Optics and Lasers in Engineering, 83, pp.48-58.

Determining prerogative by factors of significance, logical or programming controller

Ivan Didenko
Bachelor of computer science
Department information technologies of design
Sumy state university, aegjndjhf@gmail.com

***Abstract* – Advances in information technology design computer systems and networks helped to deepen the study of the processes that can lead to loss of integrity, confidentiality or availability of information or even its destruction. Information technology protect the confidentiality of information in computer systems and networks for the present time by following directions - SDN (software-defined networking) and ACI (Application Centric Infrastructure). Given the trends of information technology and approaches to problem-solving information security important task is to determine the appropriate direction through the use of a decision support system to further improve the selected information technology. Using decision support system conducted experimental verification of analytical model that allows you to answer about the merits of one or another system chosen key measurements and metrics, identify the strengths and weaknesses of the systems that will allow for further development and open up new opportunities.**

Keywords: information technology, information security, controller, infrastructure, decision support system.

I. INTRODUCTION

The evolution of mankind is proportional to the development of technologies used for human development. Society has allocated a priority of human qualities such as communicability, mobility, ability to learning. All of these qualities depend of the ability to exchange information.

The procedure for data transfer protocol transport layer, acquires many sets of solutions for information security against unauthorized use. Required complexity of software and hardware components it leads to an increase in cost, power consumption, size, complexity management, and sometimes incompatible.

The choice of information technology options that provide protection of confidential data while refuted an integral component of information security regarding quantity, quality and price parameters, set the direction for scientific research on improving the organization of information security without creating new data processing tools.

Using decision support systems the process of analyzing the submitted data which requires determining their impact systematically, but available at this time, the possibility of specific software and existing computing power not abolished following a scientific approach and preserve critical thinking. Understanding the processing of mechanics and transmitting information, above implementations, trends and approaches, historical experience and the development of information technology, permitting ask relevant queries and give them a correct assessment creating the correct numerical approximate calculation. Stabilizing factor influencing the development and competitiveness defined set of technologies considered certain set of priorities. Using decision support systems conducted experimental verification of analytical model allows to evaluate the advantages and disadvantages existing technologies to protect data confidentiality protocols in the transport layer on key performance factors and depending from demand, Also identify the strengths and weaknesses of these technologies, will justify the need for further development and modernization of the prerogative.

II. DEVELOPMENT MODEL DATA

Stabilizing factor influencing the development and competitiveness defined processing technologies and information security is considered a set of specific sets of priorities are reflected in a general model of figure 1.

Logically highlight the system's ability to function without changing its own structure, and stay in balance, the need to rationalize the industry in which information technology is fully functioning and inseparable from it released a set of essential features, characteristics and properties that distinguish a functioning information technology from others and give it the certainty of information technology-related factors as a whole. As a definition based on measurements of the move process conditions for the physical implementation of information technology introduced the concept of measuring the time interval.

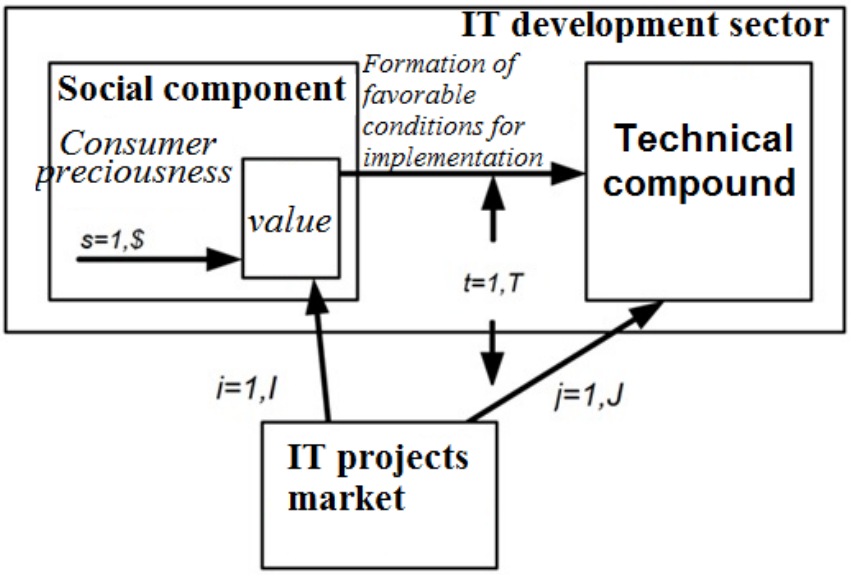


Figure 1. Analytical model

The internal stabilizing factors - physical, property product, its ability to meet specific needs, determined by the degree of usefulness of the product and the level of expenditure required to obtain.

Based on data published by a consortium ISTF (Internet Security Task Force) [1], and the Computer Security Institute (San Francisco, CA, USA) [2], that the information security issues highlighted the most important factors impressions systems to protect information from existing threats:

- social (human);
- technical (physical).

Dedicated components of the problems that need further research to determine the value of the prerogatives of development, modernization and implementation to further placing on the market of IT products need to study relevance towards the development of information technology to protect information transmitted on the low Internet protocols.

Identifying factors and indicators dependencies covering the need for rationalization of the industry in which information technology is fully functioning and inseparable from it released a set of essential features, characteristics and properties that distinguish a functioning information technology from others and give it the certainty of information technology-related factors as a whole. Considering dependency and compatibility events, implementation of information technology is presented as:

$$P = \sum_{j \to 1}^{j} (j) + \sum_{i \to 1}^{i} (i) + \sum_{s \to 1}^{s} (s) \qquad (1)$$

Modern automated information protection system implemented as multi-level system of distributed computing resources for switching common network connection interfaces basic element of which is the programmable logic controller, Its function is the current processing of real-time data and support for a set of common interfaces. The cost of such programmable logic controllers sometimes appreciated is too high, but when using SDN (software-defined networking) role of physical switches and routers is not so important. Information technology protect the confidentiality of information in computer systems and networks for the present time by following directions - SDN (software-defined networking) [3], [4] and ACI (Application Centric Infrastructure) [5], [6].

Using decision support system "Choice" [7], which provides a quantitative characteristic presented for the study of alternative technologies to protect information transmitted, determined the most significant benefits of a given performance for a reasonable solution.

The calculations regarding the criteria for determining priorities established the importance of information technology benefits that will be selected for further advancement in the market of IT services as the most current product.

Setting priorities is the fundamental basis for development implemented in the form of projects for the creation, development and dissemination of technology. Determining, advantages of information technology that provides confidential data in the existing technologies to protect data confidentiality protocols in the transport layer concerning quantity and quality of information security, showing on figure 2.

The criteria can be applied to the entire spectrum of computer systems, according to established performance

criteria evaluated prioritize the importance of information technology that is compared to determine the factors of comparative advantages in scale to assess the reliability of mechanisms and processes of information security in computer systems, information security evaluation in computer systems and their suitability for processing information that needs protection.

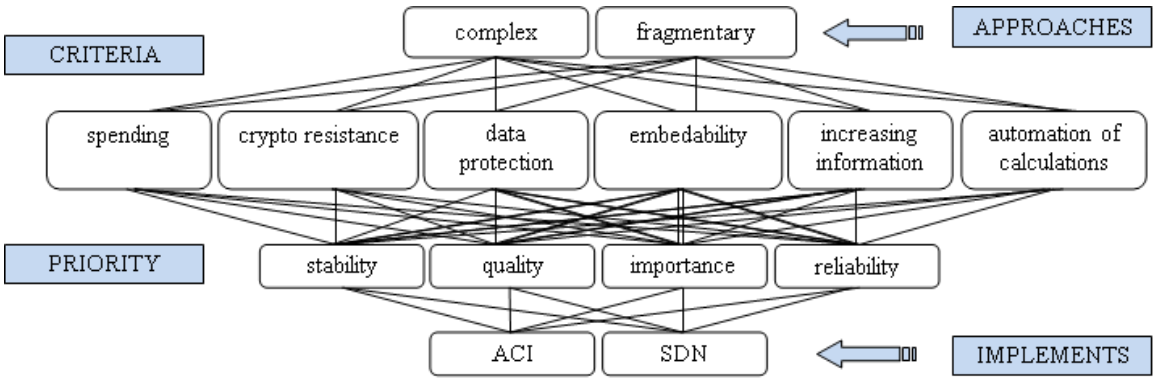


Figure 2. Model calculations

For comparative diagrams figure 3-4, the obvious benefits of information technologies protect information transmitted on specified priorities. Determining, the benefits of base (guidelines) to develop computer systems that must be implemented information security function.

Given the trends of information technology and approaches to problem-solving information security important task is to determine the appropriate direction.

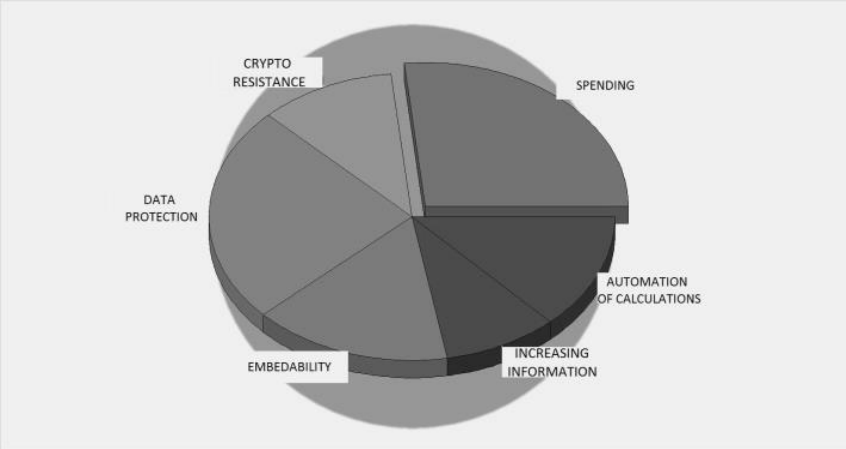


Figure 3. Comparative schedule defining the criteria of logic controller

The basic element of automated information protection system is the programmable logic controller, its function is to present data processing in real time and support a set of common interfaces. Overall programmable logic controller sold by industrial computers. The cost of these, programmable logic controller is too high.

One promising, avenue automation of information security is the implementation of programmable logic controllers based on microcontrollers, whose value is constantly decreasing simultaneously with the expansion of their functionality.

However, inexpensive microcontrollers limited computing power and support the lower set of interfaces. Therefore, in practice, implement the required interface software or hardware, which significantly affects the efficiency of information technology management.

The result of the experiment was the present of the justification for determining the importance of the prerogative, based on the separation of the comparison factors from the point of view of the importance for the further development of information technologies.

Have been investigated the determination where examined factors, which the definition in the current evaluation version information technologies low Internet protocol information transfer that protect confidential data on the evaluation version information technology providing protection of confidential data in Internet

protocols as proposed general model of prioritization of information technology that will protect the transmission of information in automated systems relative to market demand.

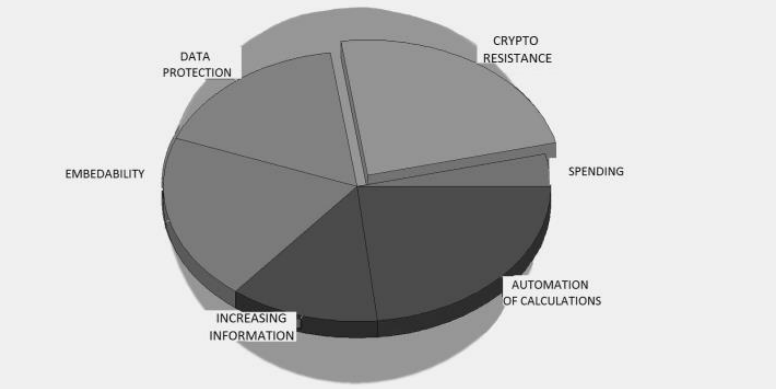


Figure 4. Comparative schedule defining the criteria of programming controller

The determination in the evaluation version information technology providing protection of confidential data in Internet protocols set by the values regarding high quality and quantity, and price performance. Priority of information technology that will protect the transmission of information in automated systems displayed in the chart figure 5.

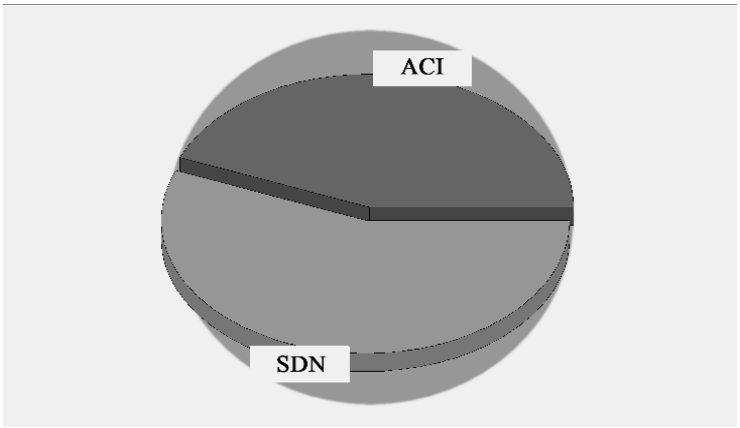


Figure 5. Selected information technology further improvement

III. CONCLUSIONS

In this scientific work the existing advanced areas of information technology network infrastructure and proposed a general model for the determination of the criteria for information technology data protection Internet protocols, low level in view of stabilizing factors and combination of certain sets priorities influence on the development and competitiveness defined processing technologies and information protection .

The result of the scientific experiment are theoretical generalization of existing approaches to ensure the development of information technology information processing data at the network level, and identified the benefits of information technology providing optimum protection of data in the Internet Protocols low.

REFERENCES:

[7] National Cyber Security Centre, Cyber Security Assessment Netherlands csan 2016. NCSC, 2016.
[8] Analytical center InfoWatch, Global investigation of leaksconfidential information in 2016. Moscow, InfoWatch, 2017.
[9] Open Networking Foundation, «Software-Defined Networking (SDN),» *opennetworking.org*, 2017. [Online]. Available: https://www.opennetworking.org/sdn-resources/sdn-definition
[10] L. Barash, «SDN: Theory and practice,» *KO IT for business*, May 23, 2015. [Online]. Available: http://ko.com.ua/sdn_teoriya_i_praktika_110584 [Accessed May 23, 2015].
[11] Cisco, «Cisco open SDN controller,» *Cisco.com* [Online]. Available: http://www.cisco.com/c/en/us/products/cloud-systems-management/open-sdn-controller/index.html
[12] S.L Blyumin and I.A. Shuykova, Models and methods adoption of solutions in terms uncertainty. Lipetsk: LEGI, 2001.
[13] S.L. Blyumin, I.A. Shuykova Models and methods adoption of solutions in terms uncertainty. - Lipetsk: LЭHY, 2001. - 138 p.

SESSION 7

- # INFORMATION TECHNOLOGIES AND LAW

International Legal Measures Against Information Warfare

Vladyslava Zavhorodnia[1], Maria Kuntsevych[2], Alla Vasylenko[3]
Sumy State University, [1]v.zavhorodnia@uabs.sumdu.edu.ua, [2]m.kuntsevich@uabs.sumdu.edu.ua

***Abstract* – The paper explores possible international legal measures against information warfare and ways of international law application to interstate informational conflicts. The authors attempt to formulate legal definition of informational warfare and identify its essential features. Two types of hostile actions on the criterion of targeting are distinguished as humanitarian and cyber forms of information warfare. The conclusion that the article's authors draw is that the contemporary international law does not establish an appropriate legal regime to information interstate conflicts. A universal international treaty is needed in order to prevent states from information aggression. The concept of its aims and main provision is also suggested.**

***Keywords* – information warfare, information aggression, cyber-attacks, international legal measures against information warfare.**

I. INTRODUCTION

In the XX and at the beginning of the XXI century creation of an alternative world picture and an alternative reality for inhabitants through informational technologies became a common practice for authoritarian and totalitarian states. The majority of those people simply consume proposed information and don't want or for various reasons aren't able to analyze various sources. Democratic states generally don't use such practices, but also makes extensive use of informational resources in political or military purposes. Thus, the use of information technologies had become a real practice of international relations, but wasn't in any way dealt with within the realm of international law.

The informational technologies as well as all achievements in the civilization of mankind could be used both for the common good and for causing harm. A new cyber weapon has appeared, capable of destruction of the whole state's informational structure. Moreover, several states (Estonia, Iran, Germany, USA and others) had already been a subject of cyber-attack. British journal «The Economist» already in the year 2010 defined cyberspace as 'the fifth domain of warfare, after land, sea, air and space'[1].

Hostile propaganda informational psychological influence on society through telecommunication technologies fully experienced by Ukraine is a real threat to sovereignty, independence and territorial integrity of the states. At the same time, legal mechanisms allowing providing legal certainty in such relations between the states, avoiding information wars and enquiring international responsibility, those are not provided by the international law. This article focuses on defining possible international legal measures against information warfare and ways of international law application to relations between states resulting from informational conflicts.

II. THE LEGAL CONCEPT OF INFORMATION WARFARE

The nature and extent of international warfare consequences had profoundly changed after appearance of Internet. Before this event false information emanating from disinformation campaigns as instruments of propaganda has used as a main tool of informational war. Internet appearance led to emergence of the notion of "cyber-attack", targeted both at information itself and a whole information system. New bases for the propaganda had become social networks and Internet media resources. While Internet becomes a space for many social activities cyber-attacks pose a great threat to the national security of each state. Computer viruses can be a tool of attacks on servers of banks, state bodies, and life support system of cities, control systems at nuclear facilities, chemical plant and other potentially dangerous objects. Thus, in modern world the informational technologies could be used as a real weapon in interstate conflicts. The main problem is to enquire international responsibility and to find the persons responsible for the hostile informational actions, because determining the occurrence of an act of the informational aggression is always difficult. Therefore, states always deny their responsibility for such hostile actions, using the advantages of lack of appropriate legal regulations.

Definition of the notion of "informational war" and making necessary international legal norms are a very difficult task, insofar as the attention of scientists recently was mainly paid to the problem of private cyber-attack avoidance [2; 3; 4]. We contain that information warfare is a state of emergency due to actions causes or capable to cause the threat to informational security and targeted both at information itself (its distortion, change or destroying) with the goal of psychological influence on inhabitants and a whole information system and information processing tools of other state in order to disrupt normal operation of informational systems and to lead difficulties in work of authorized users.

There are several essential features of information warfare:

- total impact (effect of information weapons or related technologies mostly isn't individualized; it usually targeted at info systems, thereby causing

harm to the whole country's population);

- there is a significant differences between information warfare and traditional war (enemy doesn't cross the boundary, it is difficult to prove the fact of state sovereignty violation; there is no bloodshed, but the systems operation is blocked and critical information infrastructure is disrupted or destructed);
- methods or means of information warfare are hostile, but not always unlawful (the outbreak of cyber-attack could take place in a great amount of information requests targeted at one informational system);
- tracing the origins of cyber-attacks is very difficult;
- a global character of threat (cyber-attack targeted at one state could become a serious threat to the human community, so far as the network of one state is closely linked to networks of others and its consequences almost always are unpredictable).

The main distinguishing between the information warfare and cyber-attack targeted at private networks is based on criteria of subject of hostile actions. It is possible to assimilate information warfare to a specific type of armed conflict. The role of its parties could be played by states, state-like entities and international organizations. The participants of the information warfare are individuals and could be divided into two groups: those who directly participated in hostilities (combatant) and non-combatants. Cyber-attack could be initiated by subject can not be considered as a part of international warfare. In such case hostile actions should be assimilated to cybercrime and entail criminal responsibility. Information warfare as state of emergency caused by unlawful acts of other state targeted at information security. These hostile actions (form of information warfare) could be divided into two types on the criterion of targeting: ***humanitarian*** and ***cyber***. Humanitarian forms include acts targeted at people minds, modification or distortion of informational world picture, such kind of worldviews transformation which is advantageous for belligerent in the conflict. This form of information warfare can be manifested itself mostly in hostile propaganda in order to influence psychologically on inhabitants of foreign state or to destruct unfavorable or important information. Cyber forms include acts targeted at systems of receiving, processing and distributing information. It should be mentioned that such approach is rather nominal, therefore all-out information war should be waged with coordinated measures of both types.

III. The Legal Provision of International Information Security

Global security issues had been discussed extensively at the international level in the last decade of 20th century. In 1998 the UN General Assembly adopted the UN Resolution A/RES53/70, entitled 'Developments in the field of information and telecommunications in the context of international security'. The revised resolution A/RES/54/49 of the same title was adopted in 1999. It pointed to the danger of informational threats in both civil and military fields. Subsequently, numerous resolutions had been adopted and the General Assembly remains actively seized of this matter.

Internationally legally binding instruments in the field of cybersecurity are elaborated in the framework of the Council of Europe. There are Convention on Cybercrime (2001) and Additional Protocol to the Convention on Cybercrime, concerning the criminalisation of acts of a racist and xenophobic nature committed through computer systems (2003). The Convention on Cybercrime was the first successful attempt to resolve the issues of information security. Its main objective, set out in the preamble, is to pursue a common criminal policy aimed at the protection of society against cybercrime, especially by adopting appropriate legislation and fostering international cooperation. The Treaty seeks to prevent and eliminate the crimes committed via the Internet and other computer networks, but it does not deal with the rules of international warfare.

At the national level the cyberspace policy of United States is of interest. It is based on the vision that the United States reserves the right to use all necessary means against hostile acts, including significant cyber-attacks, directed not only against the US government or military but also the economy. The significant consequences of cyber operations are: loss of life, significant responsive actions against the United States, significant damage to property, serious adverse U.S. foreign policy consequences, or serious economic impact on the United States. A retaliatory strike in response to an attack could be launched after Presidential approval [5].

The European Union has also some achievements in cybersecurity regulation. The Directive on security of network and information systems (the NIS Directive) was adopted by the European Parliament on 6 July 2016. The Directive provides network and informational strategy. It establishes a duty on Member States to adopt national provision of responding to cyber-attack and exchanging of information.

In 2016 the European Parliament adopted the Resolution on the EU strategic communication to counteract propaganda against it by third parties. The resolution stressed that the EU, Member States and citizens are under growing, systematic pressure to tackle information, disinformation and misinformation campaigns and propaganda from countries and non-state actors, such as transnational terrorist and criminal organisations in its neighbourhood. The Resolution is not a legally binding document, and does not have an enforcement mechanism. It initiates the creation of strategy to counteract anti-EU propaganda and the

adoption of measures to provide a target audience with adequate and interesting information about EU activities.

As evinced by the above overview nowadays States have not agreed on establishment of international mechanisms to counter the threats of information warfare at both universal and regional level. Moreover, there is no common approach on provision information security, restricting hostile propaganda, and prevention cyber-attack related to inter-state relations

In this context the provisions of the Tallinn Manual on the International Law Applicable to Cyber Warfare are of considerable interest. The Manual was developed by NATO's Cooperative Cyber Defense Centre of Excellence and presented in 2013. It is not a regulatory document and does not represent the official policies of NATO. This non-paper is based on existing treaties relating to the law of armed conflict, international law on State responsibility and other provisions of the international law. The Manual is an attempt to develop an international legal mechanism applicable to cyber operations, both conducted by and directed against states. It defines legal concept and types of cyber-attack, establishes criteria for distinguishing between military and nonmilitary targets, regulates the means and methods of cyber warfare. The protections of children, journalists, medical and religious personnel, UN's personnel, natural environment, cultural property, and objects indispensable to survival are also setted.

The second edition of the Tallinn Manual has been drafted in 2017. Tallinn Manual 2.0 adds a legal analysis of the more common cyber incidents that states encounter on a day-to-day basis, and that fall below the thresholds of the use of force or armed conflict. As such, the 2017 edition covers a full spectrum of international law as applicable to cyber operations, ranging from peacetime legal regimes to the law of armed conflict. The analysis of a wide array of international law principles and regimes that regulate events in cyber space includes principles of general international law, such as the sovereignty and the various bases for the exercise of jurisdiction. The law of state responsibility, which includes the legal standards for attribution, is examined at length. Additionally, numerous specialised regimes of international law, including human rights law, air and space law, the law of the sea, and diplomatic and consular law are examined within the context of cyber operations.

Despite no binding force, Tallinn Manual becomes an influential resource for legal advisers around the world. But it should be noted, that both versions on Tallinn Manual do not cover issues of hostile propaganda in cyberspace.

IV. THE PERSPECTIVES OF INTERNATIONAL LEGAL MECHANISM CREATION

Global information space should be considered as the common heritage of humankind, including the fair and equitable sharing of benefits. Therefore the information warfare should be strictly prohibited by international law. The development of international principles of behavior in cyberspace is very complicated. It should be possible to use an analogy in international law to create appropriate legal measures against information warfare. International law establishes legal regime of international armed conflicts, provides the orderly use of outer space, the high seas and other areas, concerning the national interests of all States. Therefore existing body of principles and legal standards in these fields can be applied to the problems of international information security. It should be possible to identify the definition of information aggression. As stipulated in the Charter and in United Nations resolutions the 'aggression' is use of armed force by a State against the sovereignty, territorial integrity or political independence of another State, or in any other manner inconsistent with the Charter of the United Nations. The UN's Resolution 'Definition of Aggression' mentions acts, regardless of a declaration of war, qualified as an act of aggression (Article 3). The information aggression might be defined in the similar way. Without claiming to provide a comprehensive analysis of information aggression we will try to indicate its essential features:

- it is an attack by one State against another;
- an aggressor State attacks information and communication technology systems and infrastructure, commits acts characterised as cyber and/or humanitarian type of information war;
- there is an evidence to conclude that perpetrators (hackers, dishonest journalists, bloggers, owners of 'fake' pages in social networks and others) are in any way associated with the state, and government structures are involved in cyber-attack or hostile propaganda;
- the perpetrators within the jurisdiction of State impinge on the sovereignty and information security of another State with impunity, as they have constantly been protected and sheltered from legal accountability by aggressor.

As Shackelford notes, cyber-attacks like nuclear warfare, do not discriminate between combatants and noncombatants, nor do they pass the test of proportionality. If the use of nuclear weapons is subject to the rules of the international law, so too should cyber-attacks. Nuclear weapons are not declared illegal, but methods and means of warfare which would result in unnecessary suffering to combatants, are prohibited. This principle is just as applicable to cyber war as it is to nuclear war [6]. Cyber-attacks like nuclear warfare cause mass destruction. They do not distinct military and civilian targets and can destroy objects indispensable to survival. Complete destruction of nuclear weapons is not required by the Treaty on the Nonproliferation of Nuclear Weapons. But its proliferation is prohibited. A similar approach could also be used to malicious software. It is

not practically possible to prohibit the development of it. However, it could be argued that the prevention of the proliferation of malware is the most effective manner of protecting information security. Legal provisions relating to informational warfare should be aimed at neutralizing the threats of cyber-attack. It should be possible to achieve an interstate agreement on prohibition hostile propaganda, provided that it would be adequately defined. The treaty must contain clearly established criteria and parameters set out in to identify hostile information influence. Four elements are needed in order for the acts to be qualified as hostile propaganda: firstly, systematic character of disseminating false information; secondly, the element of intent; thirdly, the specific purpose; and lastly, the involvement of a State official, at least by support or acquiescence.Although there is some resemblance between the cyberspace and outer space, both of them are incredibly vast areas of the international commons. International law does not permit outer space or cyberspace to be nationalized. Space and telecommunications systems are also intertwined, including in such functions as communications relay, imagery collection, missile warning, navigation, weather forecasting, and signals intelligence. 1967 UN Outer Space Treaty analysis allows adapting its provision to the needs of the international legal regulations on information warfare, especially regarding: the prohibition of occupation of outer space (it can be ascertained that obtaining control over information systems of the state by aggressor could be seen as an occupation of information space); freedom of exploration and use of outer space (every individual should be guaranteed the right to Internet access, understood to mean a right of unlimited access to informational resources and their using for his or her own advantage with the exception of violation of human rights or causing harm to a legally protected interests); use of outer space exclusively for peaceful purposes (use of information networks for peaceful purposes, prohibition of information aggression);international liability for damage caused by space objects (analogically state should be responsible for the damage caused by informational objects (computer programs, computer viruses etc.) those had been loaded to the network by it).

CONCLUSIONS

The certain international law provisions could be applied to cyber-attack and hostile propaganda, yet they are unable to ensure comprehensive legal measures against information warfare. The international community's efforts should focus on the conclusion of a universal international treaty and establishing of appropriate legal regime in order to prevent information warfare. The treaty should include 1) legal definitions of information warfare and information aggression; 2) prohibition of intentional hostile propaganda and using of cyber weapons; 3) responsibility of States for information aggressive acts; 4) the allocation of the burden of proof in information warfare matters; 5) the rationale for the use cyber-attack in response. Self-defense attack should be allowed when other means failed; 6) the obligation of the States to penalize intentional and/or recurrent acts of disseminating false information about another State. Otherwise, the State should be held responsible for information aggression.

Special non-governmental nonprofit organization such as The Internet Corporation for Assigned Names and Numbers (ICANN) which is responsible for coordinating the maintenance and procedures of several databases related to the namespaces of the Internet, ensuring the network's stable and secure operation, could be established for combating information warfare. This organization should take role on identification of harming activities on the Internet; take out it's an independent evaluation and block if needed. Incidental disputes and conflicts arisen as a result of blocking the activity of some users should be resolved by independent arbitration tribunal, established for such purposes. In the case of absence of reasonable suspicion that the state is involved, a case should be put on trial at the national court.

REFERENCES:

[1] *Cyberwar. The threat from the Internet*, Economist (July 1, 2010), http://www.economist.com/node/16481504?story_id=16481504&source=features_box1.

[2] O. Radutny "Criminal law measures against Internet trolling", *Internet and Law*, vol. 3(15), 2016, pp. 110-114. (in Ukrainian)

[3] O. Shyrokova-Murarash, Yu. Akchurin "Cybercrime and cyberterrorism as a danger to the informational security: international legal aspect", *Internet and Law*, vol. 3(15), 2016, pp. 76-81. (in Ukrainian)

[4] International standards of cybersafety and their appliance in Ukraine, Round Table Papers, Kharkiv: Pravo,, April, 19, 2016, 88 p. (in Ukrainian)

[5] M. Roscini *Cyber Operations as a Use of Force in International Law*, U. of Westminster School of Law Research Paper No. 16-05, 2015, p. 28.

[6] Scott J. Shackelford, *From Nuclear War to Net War: Analogizing Cyber Attacks in International Law*, 27 Berkeley J. Int'l Law. 192, 2009.

NOTES

Наукове видання

СУЧАСНІ ІНФОРМАЦІЙНІ СИСТЕМИ І ТЕХНОЛОГІЇ

Матеріали
П'тої міжнародної науково-практичної конференції
(Суми, 17 – 19 травня 2017 року)

Відповідальний за випуск В. В. Шендрик
Комп'ютерне верстання: Ю. В. Парфененко, О. В. Шулима

Формат 60×84/16. Ум. друк. арк.. 7,97. Обл.-вид. арк. 6,65.

Видавець і виготовлювач
Сумський державний університет
вул. Римського-Корсакова,2, м.Суми, 4007
Свідоцтво суб'єкта видавничої справи ДК № 3062 від 17.12.2007