Information technology Applications

Aplikácie informačných technológií

1 | 2018

PAN-EUROPEAN UNIVERSITY
Faculty of Informatics

VZDELÁVANIE VEDA VÝSKUM
EDUCATION SCIENCE RESEARCH

International Journal of Information Technology Applications (ITA)

Volume 7, Number 1, June 2018

AIMS AND SCOPE OF ITA

The primary aim of the International Journal of Information Technology Applications (ITA) is to publish high-quality papers of new development and trends, novel techniques, approaches and innovative methodologies of information technology applications in the broad areas. The International Journal of ITA is published twice a year. Each paper is refereed by two international reviewers. Accepted papers will be available online with no publication fee for authors. The journal is listed in the database of the Russian Science Citation Index (RSCI). The International Journal of ITA is being prepared for the bibliographic scientific database Scopus.

International Journal of Information Technology Applications (ITA)

Published with support from

SSKI
SLOVENSKÁ SPOLOČNOSŤ PRE KYBERNETIKU A INFORMATIKU PRI SAV
SLOVAK SOCIETY FOR CYBERNETICS AND INFORMATICS

Instructions for authors

The International Journal of Information Technology Applications is welcoming contributions related with the journal´s scope. Research papers in the range approximately 10 standard pages are reviewed by two international reviewers. Reports up to 5 standard pages and information notices in range approximately 1 standard page are accepted after the decision of editorial board. Contributions should be submitted via e-mail to the editorial office. Contributions should be typewritten in English or Russian language with extended abstract in English. Text design should preserve the layout of the Word or LaTex template file, which may be downloaded from the webpage of journal. Contributions submitted to this journal are under the author´s copyright responsibility and they are supposed not being published in the past.

Deadlines of two standard issues per year

paper submission deadline	– end of May/end of October
review deadline	– continuous process
camera ready deadline	– end of June/end of November
release date	– July/December

Editorial office address

Faculty of Informatics, Pan-European University, Tematínska 10, 851 05 Bratislava, Slovakia
juraj.stefanovic@paneurouni.com

Published by

Pan-European University, Slovakia, http://www.paneurouni.com
Civil Association EDUCATION-SCIENCE-RESEARCH, Slovakia, http://www.e-s-r.org

Electronic online version of journal

http://www.e-s-r.org
http://www.paneurouni.com/ITA

Print

Multigrafika s.r.o., Rajecká 13, 821 07 Bratislava
Delivered to the press: June 2018 (Volume 7, Number 1, 2018) Number of prints: 50

Subscription

Contact the editorial office for order and payment details. The price is 20 EUR a year.
Older print issues are available until they are in stock (15 EUR an issue).

ISSN: 2453-7497 (online)
ISSN: 1338-6468 (print version) Registration No.: EV 4528/12

Contents

Editorial

Dear authors, dear readers, let me introduce you to the publication of our 10-journal papers.

First of all, I would like to thank you for collaborating, correcting articles on the basis of opponent testimonies and accepting changes to the template editing, edited by the editorial board so that the jounal is of ever higher quality and solving the current problems associated with the use of computer science and close disciplines in a broad practice. The articles published in this issue focus on the application of IT in transport networks, urban applications and architectures, applications in medicine, the creation of modern www-pages, image processing and visualization of processes in biotechnology.

I would like to alert you that the posts to the next issue will only be accepted in English and according to a new template that is published on the journal page and is designed to meet the current requirements and trends of the publication in modern magazines.

I wish you a pleasant summer and I look forward to your other contributions.

Juraj Štefanovič
ITA Editor-in-Chief

SIMULATION OF TRAFFIC LOADING IN MESH NETWORKS

Andrey Preobrazhenskiy, Igor Lvovich, Oleg Choporov

Abstract

This paper analyzes the performance of reactive, proactive and hybrid routing protocols when used in a wireless mobile peer-to-peer network (MANET). An appropriate algorithm for evaluating the effectiveness of routing protocols is developed. The simulation tool was an OPNET network simulator (Optimized Network Engineering Tool) Modeler version 14.0. The modeling process was based on the fact that the performance of routing protocols in networks with increased scalability and mobility was studied. Two possible situations were considered: UDP traffic was used in the first one, packet delivery is not guaranteed, generating traffic was video conference traffic, in the second one TCP traffic was used, packet delivery is guaranteed, traffic generator was traffic from the FTP server. A simulation model of the video conference network was developed, with each of the nodes in the network configured to run the appropriate routing protocol. Several parameters were analyzed: the amount of lost information (bits/sec), delay (sec), network load (bit/sec), the ratio of the number of retransmitted packets to sent, the intensity of the input stream (bits/sec). Next, a multi-criteria selection of the routing protocol by the method of hierarchy analysis for the video conference was carried out. A matrix of pairwise comparisons was formed among the selected parameters for the second level of the hierarchy, which contains criteria that affect the General goal - the choice of the routing protocol. After the simulation, it was found that the least amount of information loss is characterized by the OLSR protocol, since the amount of information loss is minimized due to the fact that a subset of repeaters is used. The minimum delay we get for the TORA protocol, but this increases the load on the network, because the service information of the protocol is noticeably increased. Based on the scale of the relative importance of the criteria, as well as the results obtained on the basis of the method of analysis of hierarchies, it was found that in the analyzed model it is preferable to use the OLSR protocol as a routing protocol.

Keywords:

Computer network, traffic loading, protocol.

ACM Computing Classification System

Network protocols, Network algorithms, Network types

Introduction

This paper is devoted to the investigation of speed reactive, proactive and hybrid routing protocols in a wireless mobile peer-to-peer network (MANET), as well as the development of an algorithm for evaluating the effectiveness of routing protocols. MANET is a type of Ad-Hoc network that operates under 802.11 standard in a discrete and dispersed environment without a common control center.

The mobile peer-to-peer network (MPN) is rapidly evolving and is an important area of the wireless mobile network. MPN is combined in it wireless networks in which mobile nodes move and control the construction of routes.

In MPN network topology changes very quickly and unpredictably, each mobile node moves without a fixed access point. MPN nodes can transmit information using multiple retransmissions, and the number of intermediate nodes can vary. Nodes must support multiple routes. If the mobile nodes are within each other's radio access zone, the source node can send the message directly to the destination node, otherwise the transfer will take place through the intermediate nodes. Therefore, today, routing is an important role to ensure reliability and efficient operation in mobile wireless networks. Efficient routing management saves you the cost of building routes, which leads to improved network performance.

Mostly, the MPN is used in military communication between soldiers, aircraft, tanks, etc. in operations management in battles, emergency management teams to rescue people, in search operations of fire or police and to deploy fixed infrastructure in case of floods, fires, etc.

Mobile offices, taxis, sports stadiums, electronic payments from anywhere, voting systems, automotive computers, the education system, from the installation of virtual classrooms, conference rooms, file sharing between users, sharing game multiple gamers [1].

The main problem in MANET is to restore communication when it is lost and to build a route for the subscriber with the minimum amount of time delay for mobile nodes in chaotic motion.

The MPN will be an integral part of the next generation of networks because of its flexibility, infrastructure, ease of maintenance, automatic configuration and cost-effectiveness.

In mobile ad hoc networks, mobile nodes should communicate with each other in order to restore communications and organise dynamic topology mobility for rapid changes of the route and the restoration of relations in a wireless network.

1 Wireless network simulation based on OPNET simulation tool

As a means of simulation the network simulator OPNET (Optimized Network Engineering Tool) Modeler version 14.0 is used. It is the most widely used commercial simulator running under the Microsoft Windows operating system and includes realizaciju studied routing protocols.

This software product not only supports MANET routing, but also provides a parallel core to support increased stability and mobility in the network. Functions of intense analysis OPNET provide the best conditions for comparisons, calculations, and coordination of output.[1].

Opnet Modeler offers users a graphical environment for creating, executing, and analyzing event modeling of communication networks. This user-friendly software can be used for a large number of problems, such as typical creation and verification of communication protocol, analysis of rotocol interactions, network optimization and planning. It is also possible to carry out with the help of the package verification of the correctness of analytical models and description of the protocols [2].

Within the framework of the so-called project editor, palettes of network objects can be created, to which the user can assign various forms of connection of nodes and connections up to the puzzle-like ones. Automated generation of network topology-ring, star, random network, is also supported and reserved by utilities for imported network topologies in various formats.

Random traffic can be automatically generated from the algorithms specified by the user, as well as imported from the standard package of real line traffic formats. The results of the simulation can be analyzed, and the graphs and traffic animation will again be generated automatically. A new feature is the automatic conversion to html.

One of the advantages of creating a network model using the software is that the level of flexibility provided by the core of the simulation is the same as for the simulation, written from scratch, but the object the build environment allows the user much faster to do the development, improvement and to produce models for repeated use.

There are several editor environments - one for each object type. Organization of objects - hierarchical, network objects (models) are connected by a set of nodes and communication objects, while node objects are connected by a set of objects, such as priority modules, processor modules, transmitters and receivers. Version software for modeling the radio channel contains a model of the antenna of a radio transmitter, receiver antenna, moving objects node (including satellites).

The behavior logic of the processor and precedence constraints determines the model of the process that the user can create and modify within the process editor. In the process editor, the user can determine a process model using the combination of the algorithm of a finite state machine (finite-state machine - FSM), and operators of the programming language C/C++.

The invocation of a process model event during the simulation is controlled by the initiation of the interrupt, and each interrupt corresponds to an event that must be processed by the process model.

The basis of communication between processes is a data structure called a package. Package formats can be defined, that is, they define which fields can contain standard data types such as integers, floating-point numbers, and package pointers (this last ability allows you to encapsulate a package simulation).

Data structure, the caller information for the control interface (interface control information-ICI) can be divided between the two process models is another mechanism for interprocessor communication, it is very convenient for teams simulation and corresponds to the layered architecture of the Protocol.

A process can also dynamically spawn child processes that simplify the functional description of systems such as servers.

Several basic process models are included in the basic package, simulating popular network protocols and algorithms, such as the border gateway protocol (BGP), the Protocol of transmission control.

Internet Protocol (TCP/IP), frame relay (frame relay), Ethernet, asynchronous transfer mode (asynchronous transfer mode -ATM), and WFQ (weighted fair queuing). Basic models are useful for the rapid development of complex simulation models for common network architectures as well as for training to give an accurate functional description of the Protocol to students.

There is a possibility of support by companies and graphics (with hypertext support) of network, node or process models [3, 4].

In this paper, the simulation is based on the study of the performance of routing protocols in a network with increased scalability and mobility. Thus, two modeling scenarios were created. The first-with UDP traffic, where the delivery of packets is not guaranteed, as the generating traffic was video conference traffic, the second-with TCP traffic, where the delivery of packets is guaranteed, as a traffic generator used traffic from the ftp server [5, 6].

The simulation model of the video conference network consisted of 20 mobile units, randomly located on the territory of 1000m per 1000m. 10 nodes from which transmitted information, and 10 nodes are received.

Each node was moving randomly throughout the territory. The speed varied between 10 Mbps and 20 Mbps throughout the simulation.

The nodes were configured with the following parameters:

- Physical characteristics of 11 Mbps
- Transmitter power 0.005 W

Traffic for the model was specified explicitly, traffic from the video conference with the following parameters was used as traffic generation:

- number of frames per second: 15 fps;
- frame size: 128x240 pixels.

Each node in the network has been configured to run the appropriate routing Protocol for AODV, DSR, OLSR, TORA for each simulation. The simulation time was 600 seconds. Each node was randomly configured with a path to simulate network mobility in space throughout the simulation.

As a result of simulation for video conference traffic, we got the following numerical data (table 1).

The simulation model of the FTP server network consisted of 20 mobile nodes and one fixed node simulating the FTP server located randomly on the territory of 1000m by 1000m.

Each of the 20 nodes contacted the server to download the file with the specified characteristics. The traffic for the model was also explicitly set on the FTP server side, the FTP traffic with the following characteristics was used as traffic generation:

- file size: 1.4 Gb;
- request time between packets : 3600 s.

Table 1 - Results of the network analysis using video conferencing

	Protocols			
	AODV	DSR	OLSR	TORA
Amount of lost information (bits / sec)	96 150 000	75 990 000	35 120 000	35 875 000
Delay (sec)	0.16800	0.19800	0.30000	0.14000
Network load (bit/sec)	19 300 000	16 900 000	17 800 000	23 000 000
The ratio of the number of packets relayed to sent	0.96	0.92	1.49	0.95
The intensity of the input stream (bits/sec)	40 000	110 000	12600	6 900 000

After a packet is requested, each packet is sent according to the Gaussian distribution law, i.e. each subsequent value is independent of the previous one and this value is characterized by a request time between packets of 3600 seconds.

Table 2 - Results for the network with the FTP server

	Protocols			
	AODV	DSR	OLSR	TORA
Amount of lost information (bits / sec)	30 200	52 500	21 900	34 500
Delay (sec)	0.015000	0.019500	0.008200	0.010100
Network load (bit/sec)	6 250 000	5 900 000	6 400 000	6 200 000
The ratio of the number of packets relayed to sent	0,359	0,351	0,255	0,550
The intensity of the input stream (bits/sec)	16 100	20 000	3 750	11 950

As with the first model in the second model, each node in the network was configured to perform the appropriate routing Protocol AODV, DSR, OLSR, TORA for each simulation.

The simulation time was 600 seconds, and nodes would access the server to download the file from the fifty-fifth second after the simulation started. Each node was randomly configured with a specific path path to simulate network mobility in space throughout the simulation.

And in the first and second cases, the simulation lasted for six hundred seconds.

For the model with FTP traffic generation we got the following results (Table 2).

2 Multi-criteria routing protocol selection by hierarchy analysis method for video conference

To make a decision based on the data obtained, we will use the method of hierarchy analysis [7, 8]. The General view of the hierarchy of the decision will be as follows (Figure 1), where Ki are particular selection criteria, Aj are possible alternatives.

A1-protocol AODV; A2 – protocol DSR; A3- protocol OLSR; A4- protocol TORA.

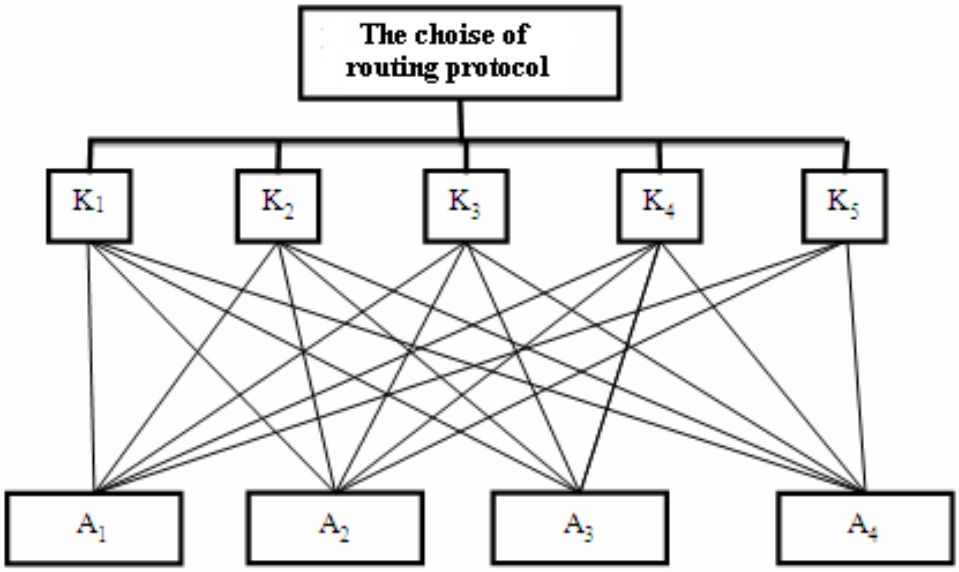


Fig.1. Hierarchy of the decision

Table 3 is a matrix of pairwise comparisons for the second level of the hierarchy, which contains the criteria that affect the overall goal - the choice of routing protocol.

Table 3 - matrix of pairwise comparisons of criteria

CRITERIA	Amount of information lost (K_1)	Delay (K_2)	Network load (K_3)	Ratio of the number of packets relayed to sent (K_4)	Intensity of the input stream (K_5)
Amount of information lost (K_1)	1	9	7	7	9
Delay (K_2)	1/9	1	1/7	1/3	3
Network load (K_3)	1/7	7	1	5	3
Ratio of the number of packets relayed to sent (K_4)	1/7	3	1/5	1	3
Intensity of the input stream (K_5)	1/9	1/3	1/3	1/3	1

Based on the results obtained in the simulation of the network with video conference traffic (table 1) and using the method of hierarchy analysis, we construct a matrix for each of the criteria.

For the criterion "The amount of lost information (K_1)»:

Table 4 - Evaluation of the importance of alternatives for the criterion of the amount of lost information (K_1)

	AODV	DSR	OLSR	TORA	Normalized estimates of priority vectors
AODV	1	0.790327613	0.365262611	0.373114925	0.130154
DSR	1.265298066	1	0.462166074	0.472101592	0.164684
OLSR	2.737756264	2.163724374	1	1.021497722	0.356331
TORA	2.680139373	2.118188153	0.978954704	1	0.348831
Sum	7.6832	6.0722	2.8064	2.8667	

For the criterion " Delay (K_2)»:

Table 5 - Assessment of the importance of alternatives for the Delay criterion (K_2)

	AODV	DSR	OLSR	TORA	Normalized estimates of priority vectors
AODV	1	1.178571429	1.785714286	0.833333333	0.277125
DSR	0.848484848	1	1.515151515	0.707070707	0.235136
OLSR	0.56	0.66	1	0.466666667	0.155190
TORA	1.2	1.414285714	2.142857143	1	0.332550
Sum	3.6085	4.2529	6.4437	3.0071	

For the criterion «Network Load (K_3) »:

Table 6 - assessment of the importance of alternatives for the network Load criterion (K_3)

	AODV	DSR	OLSR	TORA	Normalized estimates of priority vectors
AODV	1	0.875647668	0.922279793	1.191709845	0.245978
DSR	1.142011834	1	1.053254438	1.360946746	0.280909
OLSR	1.084269663	0.949438202	1	1.292134831	0.266706
TORA	0.839130435	0.734782609	0.773913043	1	0.206407
Sum	4.0654	3.5599	3.7494	4.8448	

For the criterion «Ratio of the number of retransmitted alternatives to sent alternatives (K_4)»:

Table 7 - evaluation of the importance of alternatives for the criterion of the Ratio of the number of relayed to sent alternatives (K_4)

	AODV	DSR	OLSR	TORA	Normalized estimates of priority vectors
AODV	1	0.958333333	1.552083333	0.989583333	0.270395
DSR	1.043478261	1	1.619565217	1.032608696	0.282151
OLSR	0.644295302	0.617449664	1	0.637583893	0.174214
TORA	1.010526316	0.968421053	1.568421053	1	0.273241
Sum	3.6983	3.5442	5.7401	3.6598	

For the criterion «Intensity of the input flow (K5)»:

Table 8 - Evaluation of the importance of alternatives for the input flow Rate criterion (K_5)

	AODV	DSR	OLSR	TORA	Normalized estimates of priority vectors
AODV	1	2.75	0.315	172.5	0.220069
DSR	0.363636364	1	0.114545455	62.72727273	0.080025
OLSR	3.174603175	8.73015873	1	547.6190476	0.698631
TORA	0.005797101	0.015942029	0.001826087	1	0.001276
Sum	4.5440	12.4961	1.4314	783.8463	

As a result, we obtain the values of global priorities of alternatives AODV (A_1), DSR (A_2), OLSR (A_3), TORA (A_4):

According to the obtained values, you should choose an alternative with a maximum global priority value equal to 0,324954, which corresponds to the OLSR protocol.

Table 9 - Global priorities of alternatives AODV (A_1), DSR (A_2), OLSR (A_3), TORA (A_4):

Alternatives	Criteria					Global priorities
	The amount of lost information	Delay	The network Load	The ratio of the number of packets relayed to sent	Intensity of the input stream	
	Numerical value of the priority vector					
AODV	0.130154	0.277125	0.245978	0.270395	0.220069	0.177248
DSR	0.164684	0.235136	0.280909	0.282151	0.080025	0.199034
OLSR	0.356331	0.155190	0.266706	0.174214	0.698631	0.324954
TORA	0.348831	0.332550	0.206407	0.273241	0.001276	0.298763

In the model under study, based on the results, it can be seen that the smallest amount of information loss is observed when using the OLSR protocol as a routing protocol (table 9). This is because the OLSR Protocol minimizes information loss by using a specific subset of nodes in the network, called preferred relays.

The minimum delay is achieved when using the TORA Protocol as the routing protocol (table 9), but the load on the network has increased due to a significant increase in the service information of the protocol itself.

Routing protocols AODV and DSR showed similar results. This is because both protocols are reactive routing protocols and both build on-demand route tables based on the distance vector. Their difference lies in the fact that the protocol uses DSR for routing, the routing table of the source, not the intermediate nodes.

Based on the scale of relative importance of the criteria (table 3) and the results obtained by the method of hierarchy analysis, in our model it is preferable to use the OLSR protocol as a routing protocol.

3 Multi-criteria selection of the routing protocol method of analysis of hierarchies for the FTP server

Based on the results obtained in the simulation of the network with video conference traffic (table 4) and using the method of hierarchy analysis, we construct a matrix for each of the criteria.

For the criterion «The amount of lost information (K_1)»:

Table 10 - Evaluation of the importance of alternatives for the criterion of the amount of lost information (K_1)

	AODV	DSR	OLSR	TORA	Normalized estimates of priority vectors
AODV	1	1.738410596	0.725165563	1.142384106	0.261124
DSR	0.575238095	1	0.417142857	0.657142857	0.150209
OLSR	1.378995434	2.397260274	1	1.575342466	0.360089
TORA	0.875362319	1.52173913	0.634782609	1	0.228578
Sum	3.8296	6.6574	2.7771	4.3749	

For the criterion " Delay (K_2)»:

Table 11 - Assessment of the importance of alternatives for the Delay criterion (K_2)

	AODV	DSR	OLSR	TORA	Normalized estimates of priority vectors
AODV	1	1.3	0.546666667	0.673333333	0.196709
DSR	0.769230769	1	0.420512821	0.517948718	0.151315
OLSR	1.829268293	2.37804878	1	1.231707317	0.359834
TORA	1.485148515	1.930693069	0.811881188	1	0.292142
Sum	5.0836	6.6087	2.7791	3.4230	

For the criterion «Network Load (K3) »:

Table 12 - Assessment of the importance of alternatives for the network Load criterion (K3)

	AODV	DSR	OLSR	TORA	Normalized estimates of priority vectors
AODV	1	0.944	1.024	0.992	0.247283
DSR	1.059322034	1	1.084745763	1.050847458	0.261952
OLSR	0.9765625	0.921875	1	0.96875	0.241487
TORA	1.008064516	0.951612903	1.032258065	1	0.249277
Sum	4.0439	3.8175	4.1410	4.0116	

For the criterion «Ratio Of the number of retransmitted alternatives to sent alternatives (K_4)»:

Table 13 - Evaluation of the importance of alternatives for the criterion of the Ratio of the number of relayed to sent alternatives (K_4)

	AODV	DSR	OLSR	TORA	Normalized estimates of priority vectors
AODV	1	0.977715877	0.710306407	1.532033426	0.244896
DSR	1.022792023	1	0.726495726	1.566951567	0.250478
OLSR	1.407843137	1.376470588	1	2.156862745	0.344775
TORA	0.652727273	0.638181818	0.463636364	1	0.159850
Sum	4.0834	3.9924	2.9004	6.2558	

For the criterion «Intensity of the input flow (K5)»:

Table 14 - Evaluation of the importance of alternatives for the input flow Rate criterion (K_5)

	AODV	DSR	OLSR	TORA	Normalized estimates of priority vectors
AODV	1	1.242236025	0.232919255	0.742236025	0.134307
DSR	0.805	1	0.1875	0.5975	0.108117
OLSR	4.293333333	5.333333333	1	3.186666667	0.576626
TORA	1.347280335	1.673640167	0.313807531	1	0.180950
Sum	7.4456	9.2492	1.7342	5.5264	

Table 15 - As a result, we obtain the values of global priorities of alternatives AODV (A_1), DSR (A_2), OLSR (A_3), TORA (A_4):

Alternatives	Criteria					Global priorities
	The amount of lost information	Delay	The network Load	The ratio of the number of packets relayed to sent	Intensity of the input stream	
	Numerical value of the priority vector					
AODV	0.261124	0.196709	0.247283	0.244896	0.134307	0.248582
DSR	0.150209	0.151315	0.261952	0.250478	0.108117	0.180216
OLSR	0.360089	0.359834	0.241487	0.344775	0.576626	0.343204
TORA	0.228578	0.292142	0.249277	0.159850	0.180950	0.227999

In the second case, when studying the model with the generation of ftp traffic, the smallest value of the amount of lost information is obtained when using the OLSR Protocol as a routing Protocol.

When using the OLSR Protocol, we got the lowest latency in the network, the ratio of relayed packets to sent packets, and the smaller value of the service information of the Protocol itself, but the value of the "network load" was the highest.

The largest value of the amount of lost information is obtained by using the DSR protocol. Despite the similarity of DSR with AODV, the value of this criterion is much lower when using AODV. This is due to the fact that during the generation of ftp traffic, tcp packets prevail, in which the sequence of sent data and their loss plays an important role. In the model of the network under study, high mobility of nodes prevails, which leads to intensive updating of route tables of each node. However, unlike AODV, which maintains a route table on each node, DSR is based on the source route table, which leads to an inevitable increase in the amount of information lost.

When using the TORA protocol as a routing protocol, we have similar results as with other protocols. However, the value of the ratio of the number of retransmitted packets to sent packets exceeds the values of other protocols. This is due to the hybrid nature of the protocol and the high mobility of the network under study. Since the network had high mobility and a large amount of generated traffic, the protocol had to save and use routes between all source – receiver pairs constantly, but not all routes were relevant.

As in the first model of mobile network with video conference traffic generation, in the second model it is also preferable to use OLSR protocol as a routing protocol.

Conclusion

In this paper, the problem of choosing a routing Protocol for a mobile wireless network with a cellular topology with different types of traffic is solved. Two models of wireless network were studied, one with udp traffic, where packet delivery is not guaranteed, as generating traffic was video conference traffic, the second - with tcp traffic, where packet delivery is guaranteed, as a traffic generator was used traffic from the ftp server. Four routing protocols AODV, DSR, OLSR, TORA, characterized by a number of indicators that affect the network characteristics in different ways, and the degree of influence was also different. The problem of choosing the best Protocol was related to the problem of multi-criteria selection. As a result, it was decided to use the method of hierarchy analysis as a method of selecting a suitable Protocol.

A matrix of criteria importance was compiled (table 3). The main and the most important common criterion for all protocols was the amount of user information lost, and the important criteria were the network delay and the intensity of the input stream.

The results show that the use of THE Tora hybrid Protocol in a network with traffic generated by a video conference allows to minimize the amount of lost information in conditions with an optional guarantee of package delivery. This allows it to deliver data much faster and more efficiently for applications that require high bandwidth or require little time to deliver data. However, this generates more service information for the Protocol itself.

Reactive protocols AODV and DSR showed similar results. However, under the conditions of packet delivery guarantee and high mobility, AODV Protocol is clearly superior to DSR Protocol, it has less delays and less sent service information, which leads to a reduction in the amount of lost information.

Based on the analysis of the simulation results and the use of the hierarchy analysis method to solve the multi-criteria problem of route selection in a wireless network, it can be concluded that for the considered networks the optimal choice is the OLSR routing Protocol. This Protocol belongs to the proactive protocols. It actively determines the level of network status by regularly exchanging packets between nodes in the network. The OLSR Protocol minimizes the amount of overhead by using a specific subset of nodes in the network, called repeaters, to relay control messages. This explains the low load on the network using this Protocol (table 4), which is especially effective in the transmission of information on the network, sensitive to delays.

References

[1] Qasim Nadia, Said Fatin & Aghvami Hamid Mobile Ad Hoc Networking Protocol's Evaluation through Simulation for Quality of Service // *IAENG International Journal of Computer Science*, 36:1, IJCS_36_1_10, 17 February, 2009.

[2] Wireless mesh networking. *Architectures, Protocols and Standards*. / Yan Zhang [& etc.]; ed. by Yan Zhang // Auerbach Publications, 2007.- 592 p.

[3] Alex Hinds, Michael Ngulube, Shaoying Zhu, and Hussain Al-Aqrabi A Review of Routing Protocols for Mobile Ad-Hoc NETworks (MANET) // *International Journal of Information and Education Technology*, Vol. 3, No. 1, February 2013.

[4] Mukesh Kumar, A.K.Jaiswal Vivek Kumar Singh, Rohini Saxena Performance Analysis And Energy Consumption Of Routing Protocol In Manet Using Grid Topology // *International Journal of Technical Research and Applications*, Volume 1, Issue 5 (NovDec 2013), PP. 70-75.

[5] Ermolova V. V., Preobrazhensky Yu. P. The architecture of the system exchange messages in non-routable network // *Bulletin of the Voronezh Institute of high technologies*. 2010. No. 7. P. 79-81.

[6] Preobrazhensky Yu. P. Information security - the challenges of the modern world // *Bulletin of the Voronezh Institute of high technologies*. 2017. № 2 (21). P. 60-63.

[7] Saati T. *Decision - making. Method of analysis of hierarchies* // M.: Radio and communication, 1993. - 278 p.

[8] Saati T. *Decision - making at dependences and feedbacks: Analytical networks* / edited By A. V. Andreychikova, O. N. Andreychikova. - M.: Publishing house LKI. - 360 p.

Prof. Andrey Preobrazhenskiy
Doctor of Sciences (Engineering), associate professor,
Voronezh Institute of High Technologies, Russia
E-mail: app@vivtl.ru

Prof. Igor Lvovich
Doctor of Sciences (Engineering), professor,
Voronezh Institute of High Technologies, Russia
office@vivt.ru

Prof. Oleg Choporov
Doctor of Sciences (Engineering), professor,
Voronezh Institute of High Technologies, Russia
choporov_oleg@mail.ru

INFORMATIZATION OF THE HOUSING SPHERE OF RUSSIA ON THE EXAMPLE OF ST. PETERSBURG

ИНФОРМАТИЗАЦИЯ ЖИЛИЩНОЙ СФЕРЫ РОССИИ НА ПРИМЕРЕ САНКТ-ПЕТЕРБУРГА

Sergey Kirsanov, Evgeniy Istomin, Eva Mihaliková

Abstract:

Six months ago in Russia, the state information system of housing and communal services, which allows you to check the license of any management company, communicate on the forum, find your home and information on it and about its services, get acquainted with the news in the housing sector, you can in the round-the-clock mode to use the legal framework not only for municipal issues, but also for any other areas of legislation. However, further improvement of the information system is necessary.

Keywords:

State information system of housing and communal services

ACM Computing Classification System:

Applied Computing, Computers in other domains, E-Government

Абстракт:

Полгода назад в России заработала государственная информационная система жилищно-коммунального хозяйства, которая позволяет проверить лицензию любой управляющей компании, пообщаться на форуме, найти на карте свой дом и информацию о нем и об обслуживающих его организациях, ознакомиться с новостями в жилищной сфере, можно в круглосуточном режиме пользоваться нормативно-правовой базой не только по коммунальным вопросам, но и по любым другим сферам законодательства. Однако необходимо дальнейшее совершенствование информационной системы.

Ключевые слова:

государственная информационная система жилищно-коммунального хозяйства

Введение

Сегодня в системе ГИС ЖКХ *РФ* почти 15,8 миллиона многоквартирных домов РФ (97,3%) и индивидуальных жилых домов (83%).

По состоянию на 1 октября 2017 года в ГИС ЖКХ зарегистрировано более 82,5 тысячи организаций - это все управляющие компании (УК) страны, 93% товариществ собственников жилья (ТСЖ) и жилищно-строительных кооперативов (ЖСК), 97% ресурсоснабжающих организаций.

В системе ГИС ЖКХ России размещена информация о 71,3 миллиона лицевых счетов, 95,6 тысячи тарифов, 340 тысяч проведенных контролирующими органами проверок, более 32,2 миллиона приборов учета.

С ГИС ЖКХ интегрированы 2464 ИТ-системы. Все кредитные организации также интегрированы с ГИС ЖКХ и в ежедневном режиме передают данные об оплате жилищно-коммунальных услуг (передано более 820 миллионов фактов оплаты).[1]

1 Внедрение Государственной информационной системы коммунального хозяйства в Санкт-Петербурге

В апреле 2017г постановлением Правительства СПб №270 была создана "Региональная государственная информационная система жилищно-коммунального хозяйства".[2] Целями создания региональной ГИС являются: представление информации, а также информации, обязательное размещение которой в государственной информационной системе жилищно-коммунального хозяйства поставщиками информации, осуществляющими деятельность на территории Санкт-Петербурга, предусмотрено законодательством; повышение эффективности государственного управления ЖКХ; а также обеспечение размещения в ГИС ЖКХ предусмотренной Федеральным законом информации. Оператор ГИС ЖКХ – Жилищный комитет.

Основными функциями рассматриваемой ГИС являются сбор, обработка, анализ и хранение информации; доступ к информации, содержащейся в Системе, размещение и представление такой информации в электронной форме; представление пользователям информации достоверной и актуальной информации; взаимодействие с ГИС ЖКХ, государственными информационными системами Санкт-Петербурга и базами данных, и др.

В состав Системы входят следующие подсистемы:

- подсистема "Отраслевое хранилище данных жилищно-коммунального хозяйства Санкт-Петербурга", предназначенная для сбора, обработки и хранения информации;
- подсистема "Отраслевой портал жилищно-коммунального хозяйства Санкт-Петербурга" (далее - Отраслевой портал), предназначенная для размещения в информационно-телекоммуникационной сети "Интернет" по адресу: www.gilkom-complex.ru общедоступной информации о состоянии жилищно-коммунального хозяйства Санкт-Петербурга; а также для предоставления возможностей по работе пользователей Системы в личных кабинетах;

[1] *Шадрина Т. Коммуналка уходит в "электронку". Дан старт полноценной работе ГИС ЖКХ/ Российская газета - Федеральный выпуск №7392 (226) от 5.10.2017*

[2] *Постановление Правительства СПб от 14 апреля 2017 года N 270 «О государственной информационной системе Санкт-Петербурга "Региональная государственная информационная система жилищно-коммунального хозяйства"*

- подсистема "Информационное взаимодействие с государственными информационными системами Санкт-Петербурга и базами данных", предназначенная для организации взаимодействия с иными информационными системами, в том числе с ГИС ЖКХ;

- подсистема "Система защиты информации", предназначенная для обеспечения защиты информации, содержащейся в Системе, от неправомерного доступа, уничтожения, модифицирования, блокирования, копирования, представления, распространения, а также от иных неправомерных действий в отношении такой информации; а также для соблюдения конфиденциальности информации ограниченного доступа.

Общедоступная информация, содержащаяся в данной ГИС, размещается в сети Интернет на Отраслевом портале. Доступ к указанной информации не требует авторизации пользователей информации.

Рассмотрим подробнее Отраслевой портал жилищно-коммунального хозяйства Санкт-Петербурга, который создан для обеспечения технологической возможности исполнения действующего законодательства в сфере ЖКХ всеми участниками отрасли, межведомственного и внутриведомственного электронного взаимодействия, консолидации информации жилищной сферы Санкт-Петербурга, а также для публикации в открытом доступе информации о состоянии дел в сфере ЖКХ города.

Портал ЖКХ предназначен для совершенствования информационного обеспечения населения и исполнительных органов государственной власти Санкт-Петербурга в сфере жилищно-коммунального хозяйства посредством создания механизмов сбора, обработки и предоставления специализированного доступа к информационным ресурсам исполнительных органов государственной власти (далее - ИОГВ), подведомственных учреждений, управляющих компаний, ТСЖ (товариществ собственников жилья) и других участников отрасли.

Портал ЖКХ обеспечивает единую программно-информационную инфраструктуру в сфере ЖКХ, создаваемую через интеграцию заинтересованных участников информационного обмена в единую информационную систему, интегратором и регулятором в которой выступает Жилищный комитет.

Портал включает открытую и закрытую части. Открытая информационная часть доступна без регистрации, содержит всю актуальную информацию с сайта Жилищного комитета.

Закрытая часть в свою очередь доступна авторизованным пользователям.

Для получения доступа к информации пользователю потребуется пройти процедуру регистрации. Управляющие организации, ресурсоснабжающие организации и органы исполнительной власти Санкт-Петербурга получают доступ после направления заявки на подключение к Порталу официальным письмом в Жилищный комитет, физические лица получают доступ после заполнения информации в разделе "Регистрация".

Сведения по домам и техническая часть электронного паспорта многоквартирного дома, не содержащие персональных данных граждан, доступны всем зарегистрированным пользователям. Для того чтобы получить доступ к лицевому счету пользователю потребуется ввести дополнительную информацию, подтверждающую его право доступа к ней. В настоящее время ведется работа по подключению к Порталу биллинговых систем госучреждения "Вычислительный центр коллективного пользования жилищного хозяйства" и отладка процесса обмена информацией. Функция получения информации о состоянии индивидуального лицевого счета пока неактивна. Со временем на Портале будет обеспечена возможность получения данных о своем лицевом счете для жителей всех многоквартирных домов Санкт-Петербурга.

Подсистемы Портала ЖКХ и входящие в их состав программные компоненты:

1. подсистема отображения и ввода информации конечными пользователями состоит из следующих программных компонентов:

- программный компонент «Управление и кэширование данных»;
- программный компонент «Веб-портал»;
- программный компонент «Ввод и публикация ЭП МКД для ИОГВ»;
- программный компонент «Ввод и публикация ведомостей ИЖФ»;
- программный компонент «Ввод и публикация ведомостей МОС»;
- программный компонент «Ввод и публикация информации о деятельности управляющих организаций, ТСЖ, ЖСК»;
- программный компонент «Взаимодействие пользователей Портала ЖКХ с использованием механизма сообщений»;
- программный компонент «Публикация сведений РСКР»;
- программный компонент «Публикация сведений проекта РПКРОИ»;
- программный компонент «Публикация сведений из биллинговых систем ГУП ВЦКП «ЖХ».

2. подсистема консолидации, хранения и предоставления контента (ПКХПК) состоит из следующих программных компонентов:

- программный компонент «Хранение данных»;
- программный компонент «Консолидация данных, полученных из внешних источников»;
- программный компонент «Веб-сервисы взаимодействия с потребителями данных»;
- программный компонент «Управление технологическим процессом»;
- программный компонент «Сбор и предоставление ЭП МКД для ИОГВ»;
- программный компонент «Сбор и предоставление ведомостей ИЖФ»;
- программный компонент «Сбор и предоставление ведомостей МОС»;
- программный компонент «Сбор и предоставление информации о деятельности управляющих организаций, ТСЖ, ЖСК»;
- электронный сервис «Предоставление информации о деятельности управляющих организаций, ТСЖ, ЖСК» в ГИС СПБ «СМЭВ СПБ»;
- программный компонент «Сбор и предоставление сведений проекта РПКРОИ».

Пользователями Портала ЖКХ являются:

- ИОГВ (Жилищного комитета, Администраций районов СПб и другие);
- УК (ТСЖ, ЖК, ЖСК);
- собственники, наниматели и арендаторы помещений;
- население Санкт-Петербурга.

В целях осуществления своих функций, пользователи взаимодействуют со следующими программными компонентами портала:

- Программный компонент «Веб-портал»;
- Программные компоненты «Ввод и публикация ЭП МКД для ИОГВ», программный компонент «Сбор и предоставление ЭП МКД для ИОГВ»;
- Программные компоненты «Ввод и публикация ведомостей ИЖФ», программный компонент «Сбор и предоставление ведомостей ИЖФ»;
- Программные компоненты «Ввод и публикация ведомостей МОС», программный компонент «Сбор и предоставление ведомостей МОС»;

- Программные компоненты «Ввод и публикация информации о деятельности управляющих организаций, ТСЖ, ЖСК», программный компонент «Сбор и предоставление информации о деятельности управляющих организаций, ТСЖ, ЖСК», программный компонент «Предоставление информации о деятельности управляющих организаций, ТСЖ, ЖСК» в ГИС СПБ «СМЭВ СПБ»;
- Пользователи: население СПб; УК (ТСЖ, ЖСК), Администрации районов СПб, ЖК;
- Программный компонент «Публикация сведений РСКР»
- Программные компоненты «Публикация сведений проекта РПКРОИ», программный компонент «Сбор и предоставление сведений проекта РПКРОИ»

Таким образом, можно выделить два основных компонента: front-end и back-end составляющие. На стороне front-end реализован пользовательский интерфейс, а back-end уже представлен сервером обработки запросов и СУБД.

2 Совершенствование информационной системы

В процессе функционирования Портал ЖКХ принимает информацию от информационных систем сферы ЖКХ, обеспечивает ввод информации, не содержащейся в других информационных системах, и передачу информации во внешние информационные системы и ее публикацию различным категориям пользователей.

Разработка единого API[3] для взаимодействия с другими ГИС

В текущей версии портала реализация информационного обмена осуществляется при помощи разных СУБД, что влечет проблемы совместимости. На схеме внешних информационных связей на рисунке 1 показано многообразие входных и выходных данных.

Для обеспечения совместимости приходиться разрабатывать различные средства преобразования и конвертации данных из-за разнообразия СУБД.

В этом случае возникает потребность оптимизации процесса обмена данными для унификации и уменьшении времени передачи информации.

Одним из способов для достижения этой цели является создание API для всех участников обмена. Преимуществами является то, что API определяет функциональность, которую предоставляет программа (модуль, библиотека), при этом API позволяет абстрагироваться от того, как именно эта функциональность реализована.

Если программу (модуль, библиотеку) рассматривать как черный ящик, то API — это множество «ручек», которые доступны пользователю данного ящика и которые он может вертеть и дергать.

Программные компоненты взаимодействуют друг с другом посредством API. При этом обычно компоненты образуют иерархию — высокоуровневые компоненты используют API низкоуровневых компонентов, а те, в свою очередь, используют API еще более низкоуровневых компонентов.

По такому принципу построены протоколы передачи данных по сети Интернет. Стандартный стек протоколов по сетевой модели OSI содержит 7 уровней (от физического уровня передачи бит до уровня протоколов приложений, подобных протоколам HTTP и IMAP).

[3] API — набор готовых классов, процедур, функций, структур и констант, предоставляемых приложением (библиотекой, сервисом) или операционной системой для использования во внешних программных продуктах. Используется программистами при написании всевозможных приложений.

Каждый уровень пользуется функциональностью предыдущего («нижележащего») уровня передачи данных и, в свою очередь, предоставляет нужную функциональность следующему («вышележащему») уровню.[4]

Схема нового взаимодействия с исследуемой государственной информационной системой приведена ниже на рисунке 2. Как можно наблюдать, все потоки информации проходят через API. При этом если нужно масштабировать систему, то не нужно заново разрабатывать интерфейсы обмена с различными системами, если заранее прописать этот механизм в нормативном документе.

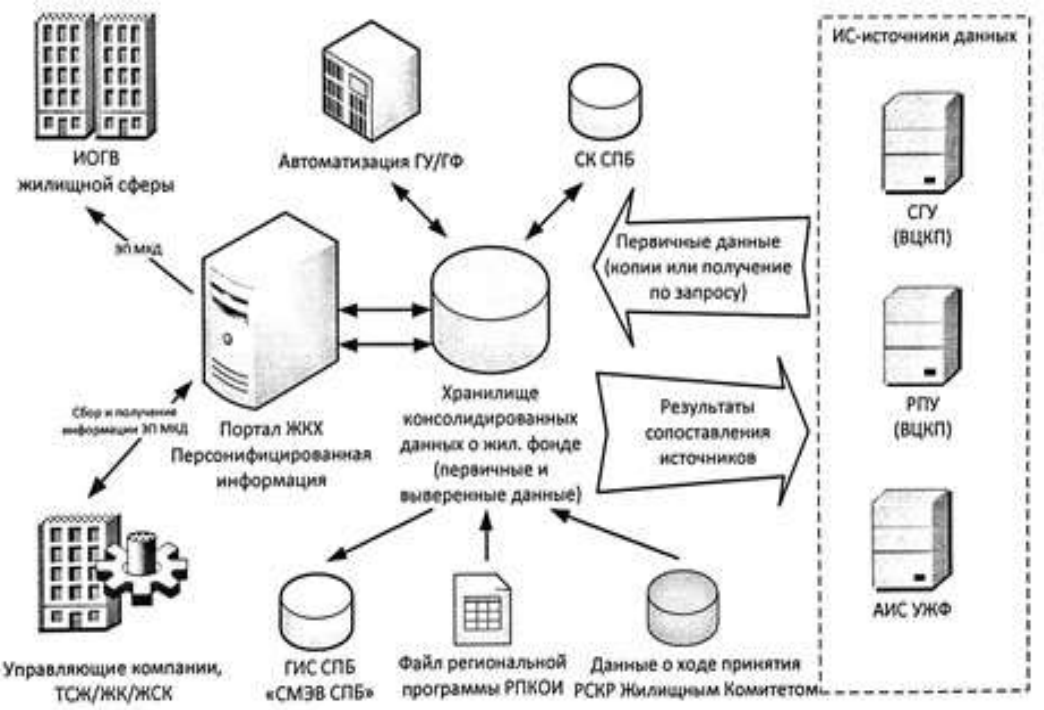


Рисунок 1. Схема внешних информационных связей

Можно составить базовые требования для реализации данного подхода информационного обмена:

Интерфейсная часть портала

Интерфейс сайта представляет собой визуальное его отображение на экране монитора компьютера и является одним из наиболее важных элементов не только интернет-портала, но и любого приложения, программного обеспечения. Именно пользовательский интерфейс отвечает за то, насколько удобно пользователю будет взаимодействовать с сайтом или приложением, соответственно, захочет ли он использовать этот ресурс и в дальнейшем.

[4] *https://ru.wikipedia.org/wiki/API*

Страница сайта оптимизирована только для экранов мониторов персональных компьютеров, что в свою очередь сужает аудиторию пользователей данным порталом. На рисунке 3 видно страницу, размещенную на всем экране монитора. Все элементы расположены относительно центра экрана. Верхнее меню – это «Навигатор по государственным сайтам Санкт – Петербурга». Он является обязательном элементом на всех государственных сайтах. Навигатор, в отличие от самого портала, масштабируется и адаптируется в зависимости от размера окна или мобильного устройства.

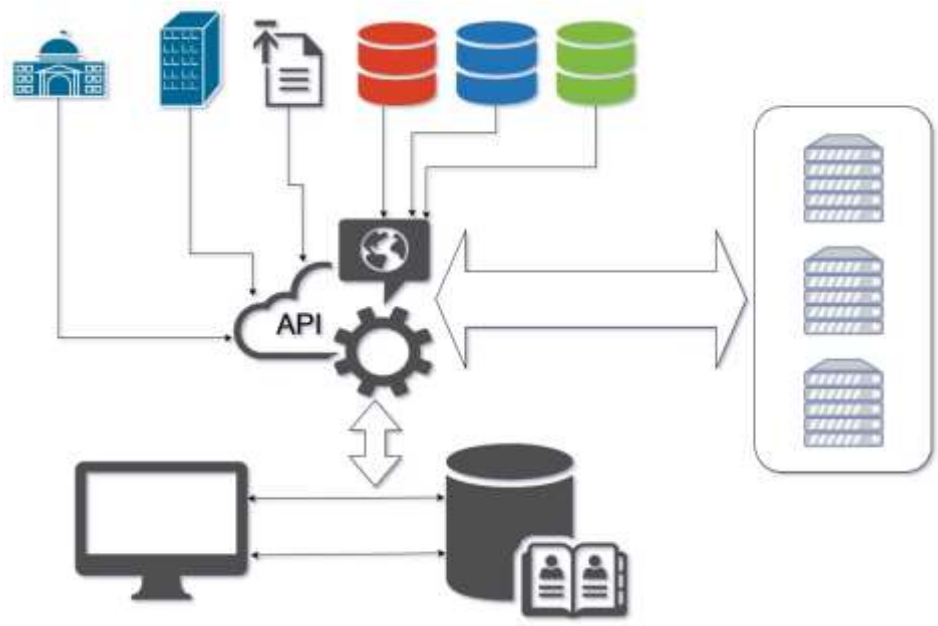


Рисунок 2. Новая схема связей с учетом нововведения

Другая недоработка главной страницы портала ЖКХ, на наш взгляд, это множество расположенных баннеров внизу. На рисунке 4 приведен пример расположения рекламных объявлений.

Для того чтобы решить проблему избыточности места, отведенного под рекламные объявление, стоит использовать горизонтальный «ролл» с баннерами, что уменьшит занимаемую площадь и улучшит «юзабилити» портала Жилищного комитета.

Следующая недоработка заключается в том, что при уменьшении ширины верстки страницы, элементы остаются на фиксированных местах не оптимизируюсь под меняющийся размер окна. Это хорошо видно на рисунке 5.

Следует оптимизировать разметку страницы и сделать ее динамически зависимой от размера окна. Это также скажется на охвате пользователей данным государственным порталом.

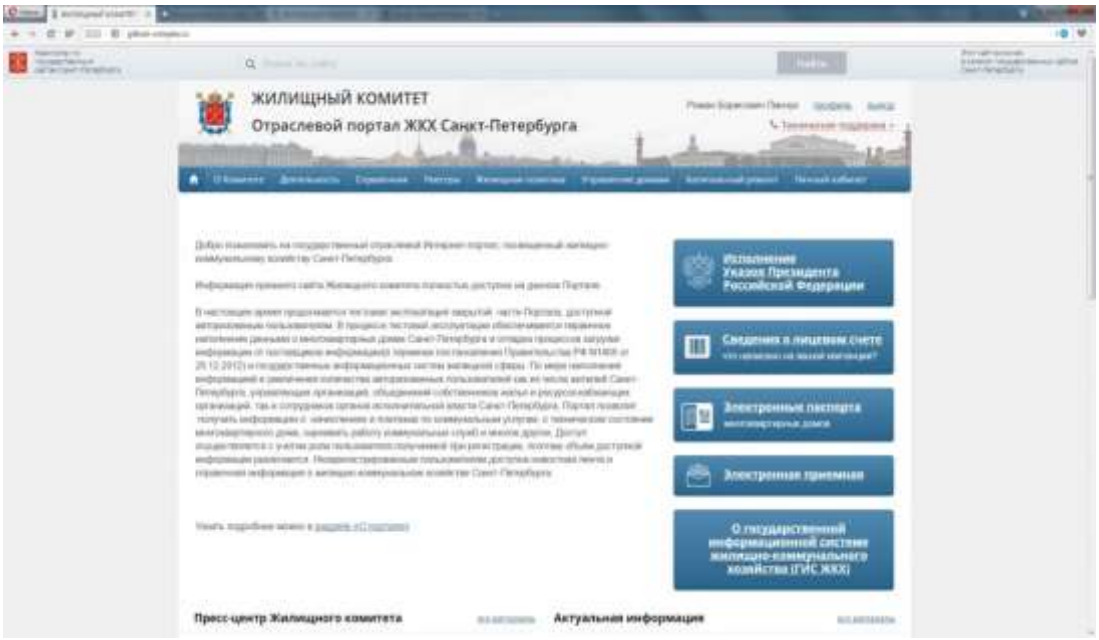


Рисунок 3. Страница портала на всю ширину окна

Рисунок 4. Пример расположения баннеров.

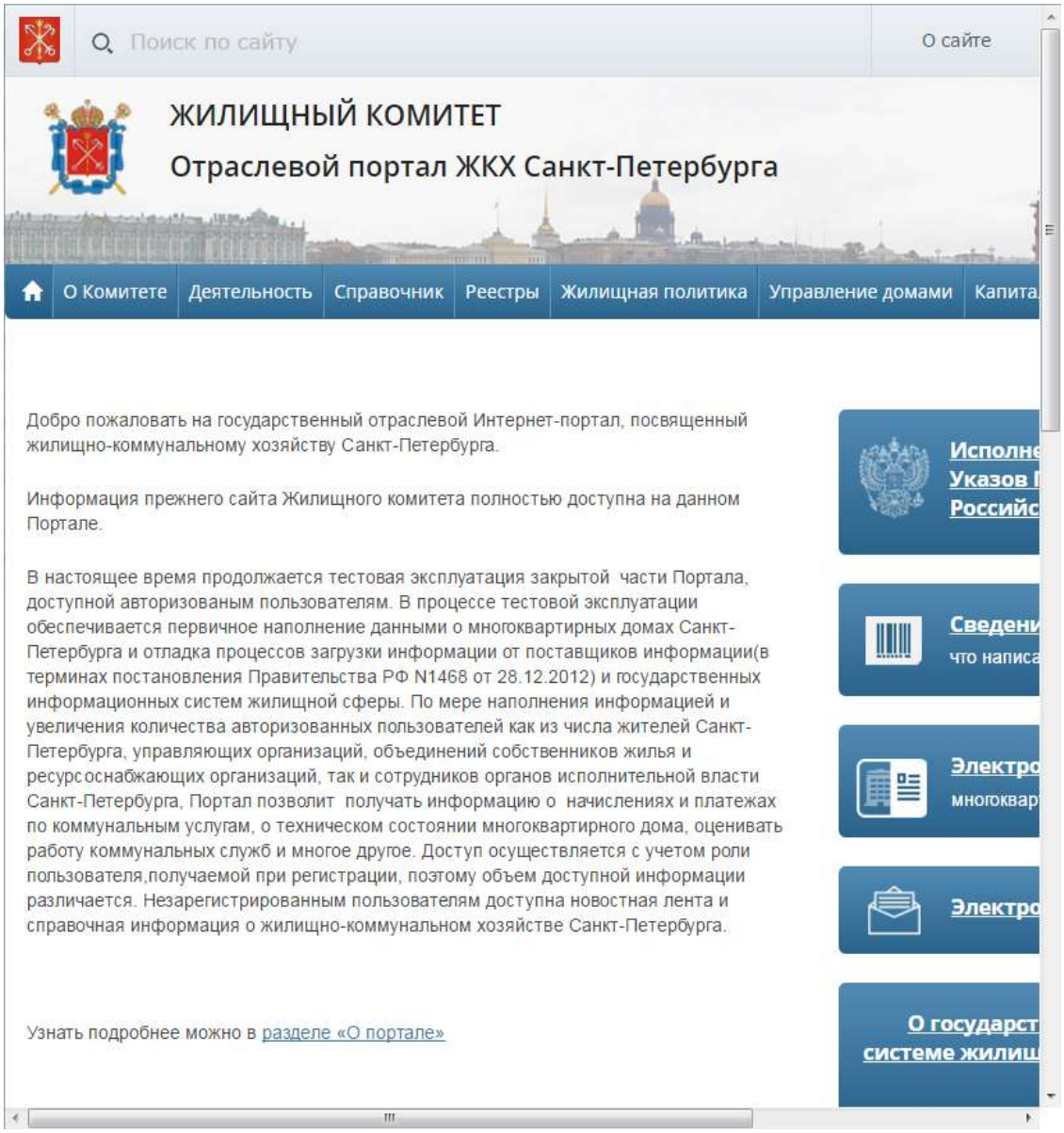


Рисунок 5. Статичное расположение элементов на странице

При исследовании портала было обнаружено несоответствие полей представленным данным на форме по запросу пользователя. Это показано на рисунке 6.

Для обеспечения правильного соответствия нужно обеспечивать проверку на разных уровнях выполнения запросов пользователей. Так как это является некорректным опубликованием информации и может привести к проблемам у оператора информации данного портала.

Краткосрочный план реализации региональной программы капитального ремонта общего имущества в многоквартирных домах:

г.Санкт-Петербург, Таллинская улица, дом 17, литера А

Год: **2016**

Основание: **В связи с утверждением постановления Правительства Санкт-Петербурга от 17.11.2016 № 1050 «О внесении изменений в постановления Правительства Санкт-Петербурга от 18.02.2014 № 84, от 18.12.2015 № 1154»**

Тип многоквартирного дома: **Дома постройки 1957-1970 гг. категории "Хрущевки панельные"**

Год ввода многоквартирного дома в эксплуатацию: **1960**

Год последнего комплексного капитального ремонта (реконструкции): **–**

Общая площадь многоквартирного дома: **3 953.60 м²**

В том числе площадь жилых и нежилых помещений: **3 595.50 м²**

Дата приватизации первого жилого помещения: **14.01.1993**

Стоимость капитального ремонта: **1 159 012.81 руб.**

В том числе за счет средств государственной поддержки: **1 159 012.81 руб.**

В том числе за счет средств фонда капитального ремонта: **0.00 руб.**

Год проведения капитального ремонта общего имущества: **2016**

Способ формирования фонда капитального ремонта общего имущества: **Счет регионального оператора**

Виды работ:

Период ремонта крыши: **1 159 012.81 руб.**

Рисунок 6. Несоответствие полей представленной информации

Заключение

Таким образом, можно сделать вывод, что определённый технологический и организационный опыт внедрения информационных систем, поддерживающих деятельность Отраслевого портала имеется, необходимо совершенствовать эту деятельность.

Анализ используемых систем автоматизации государственной деятельности демонстрирует разобщенность систем в плане стандартов на функциональность и на технологию. В существующих условиях затруднительно эффективно реализовывать федеральные целевые программы по автоматизации деятельности органов власти, поскольку отсутствует единая система стандартов и регламентов, регулирующая данную область.

Узкая целевая направленность существующих систем, как правило, диктует низкую степень поддержки межфункционального взаимодействия ведомств. Требуется создание специализированных систем для координации данного типа взаимодействия. В статье предложено решение в виде API централизованного взаимодействия подведомственных и межведомственных систем.

Необходимо создание единой системы регламентов и стандартов, которые будут иметь определяющее значение при выборе технологий, приложений и методов в проектах автоматизации, поскольку неконтролируемое развитие локально внедренных систем неизбежно приведет к возникновению конфликтов в информационной и функциональной инфраструктуре РФ.

Литература

[1] Федеральный закон от 27.07.2006 №149-ФЗ (ред. от 19.12.2016) "Об информации, информационных технологиях и о защите информации" (с изм. и доп., вступ. в силу с 01.01.2017)
[2] Федеральный закон от 21.07.2014 №209-ФЗ "О государственной информационной системе жилищно-коммунального хозяйства"
[3] Постановление Правительства Российской Федерации от 06.07.2015 N 676 "О требованиях к порядку создания, развития, ввода в эксплуатацию, эксплуатации и вывода из эксплуатации государственных информационных систем и дальнейшего хранения содержащейся в их базах данных информации"
[4] Постановление Правительства Российской Федерации от 26 марта 2016 г. N 236 г. "О требованиях к предоставлению в электронной форме государственных и муниципальных услуг"
[5] Постановление Правительства СПб от 14 апреля 2017 года N 270 «О государственной информационной системе Санкт-Петербурга "Региональная государственная информационная система жилищно-коммунального хозяйства"
[6] Приказ Минкомсвязи России № 264 Приказ от 14 июня 2016 «О вводе в эксплуатацию государственной информационной системы жилищно-коммунального хозяйства»
[7] Единый портал государственных услуг. [Электронный ресурс] – Режим доступа: http://www.gosuslugi.ru
[8] Портал Административной реформы в Российской Федерации о совершенствовании государственного управления. [Электронный ресурс] – Режим доступа: http://www.ar.gov.ru/
[9] Сайт Федеральной службы по надзору в сфере связи, информационных технологий и массовых коммуникаций. [Электронный ресурс] – Режим доступа: https://15.rkn.gov.ru/law/p8182/
[10] Сайт портала Жилищного комитета Санкт-Петербурга. [Электронный ресурс] – Режим доступа: http://www.gilkom-complex.ru/
[11] Шадрина Т. Коммуналка уходит в "электронку". Дан старт полноценной работе ГИС ЖКХ/ Российская газета - Федеральный выпуск №7392 (226) от 5.10.2017

Prof. Sergey Kirsanov
Professor, candidate of economic Sciences
Russian State Hydrometeorological University,
St. Petersburg, Malookhtinsky PR., 98, k.428
ksaimr@mail.ru

Prof. Evgeniy Istomin
Dean, Professor, doctor of technical Sciences
Russian State Hydrometeorological University,
St. Petersburg, Malookhtinsky PR., 98, k.428

Ing. Eva Mihaliková, PhD.
Pavol Jozef Šafárik University in Kosice,
Faculty of Public Administration,
Popradská 66, 041 32 Košice, Slovakia.
eva.mihalikova1@upjs.sk

SOFTWARE DEVELOPMENT FOR THE DESIGN OF STRETCH CEILINGS

Yakov Lvovich, Alexander Plotnikov, Juraj Štefanovič

Abstract

Work on the installation of suspended ceilings requires very accurate calculations and precise cutting. When creating the appropriate design systems are based on geometric calculations and key formulas describing the surface. The aim of the work was to develop a subsystem for the design of stretch ceilings, which would minimize the time and material costs that inevitably arise during the organization of production, installation of this type of structures. The classification of approaches related to geometric modeling is carried out. There are two main approaches to geometric modeling of design objects: the approach of structural geometry and the boundary approach. Constructive geometry is to create a library of geometric primitives (elementary objects), on the basis of which by means of regularized Boolean operations (intersection, Union, etc.) is to build a model. The boundary approach is based on the capabilities of piecewise analytical description of the object, description of its boundary elements (faces, edges, vertices of features and contours, nodal point of plane objects) based on algebraic equations. Geometric models have a hierarchical structure that results from the bottom-up construction. An element of a geometric structure can be: a coordinate system, a point, a curve (flat, spatial), a surface, a geometric body. For the two-dimensional case, the objects of modeling are planar contours, which consist of segments of planar curves. The elements of the topological structure of the contours are node points, line segments. The curve can be represented by a set of points, provided that they are close enough to each other. In the scatter method of geometric simulation model is used, the basis of which is the point. The use of mathematical description of the curve complicates the model, but has a number of advantages, for example, there is no need for an interpolation scheme to find intermediate points. The program is designed "Europalace" allows you to quickly build a shape with the specified parameters. There are tools to rotate the selected objects and enable the step-by-step construction mode in order to carefully consider the geometric modeling process and adjust it if necessary. The created system of design of stretch ceilings has the following features: execution of construction on the basis of the table of input data with coordinates and angles; execution of construction of an ellipse; execution of construction of a half-ellipse; execution of construction of a rectangle; execution of construction of a triangle; functions of rotation of the selected edges;
feature construction steps; the function print project.

Keywords:

Design, stretch ceilings, program.

ACM Computing Classification System

Physical sciences and engineering, Technologies, Management of computing and information systems

Introduction

The work on the installation of suspended ceilings requires very accurate calculations and no less accurate cutting.

And since in the modern world it is very difficult to imagine work without the use of information technology, when installing tension structures can not do without the help of computer programs.

The creation of such a program will significantly simplify the process of installation of ceiling structures, to make competent calculations and cutting, quickly, efficiently and without errors.

The developed system is based on geometric calculations and fundamental formulas, on the basis of which the calculation of the project of the stretch ceiling is made. On the basis of these calculations, experts will be able to mount the ceiling, accurately calculating the amount of material that will be required, and the customer will be able to see the ceiling, how it will look realized desires.

Today stretch ceiling is a popular way of finishing the ceiling. In the Russian Federation today is the development of systems of suspended ceilings, and firms that perform this kind of work, not so much - it is possible to occupy a niche in the market of companies that perform such orders.

The software that helps designers in their work is also a little bit, and these programs miss some points that could help the designer in a particular situation, so I consider the theme chosen for the final qualifying work to be reasonable and in demand, and the product developed is able to compete with existing software products of this line.

The purpose of this paper was to develop a subsystem for the design of stretch ceilings, which would minimize the time and material costs that inevitably arise during the organization of production, installation and sales of this type of structures.

1 Geometric modeling methodology

When creating CAD, we can not do without methods of geometric modeling of objects in two-dimensional and three-dimensional space.

Geometric modeling is a method of studying objects by constructing their geometric models. A geometric model of an object is a system of geometric relations describing the object. The geometric model of the object O is a triple

$$(O)=\{A, F, R\}, \quad (1)$$

where A is the set of geometric elements;
F - set of geometric attributes(characteristics, parameters);
R-set of relationships between elements and attributes.

In this model, A is the mapping of the geometric elements of object A, F is the mapping of the attributes, and R is the mapping of the relationships between the geometric elements and the attributes of object A.

There are two main approaches to geometric modeling of design objects: the approach of structural geometry and the boundary approach [1] (Fig. 1). The approach of structural geometry is to create a library of geometric primitives (elementary objects), on the basis of which the regularized Boolean operations [3, 4] (intersection, Union, etc.) are used to build the model.

The boundary approach is based on the possibility of piecewise analytical description of the object, i.e. description of its boundary elements (faces, edges, vertices of features and contours, nodal points of plane objects) by algebraic equations.

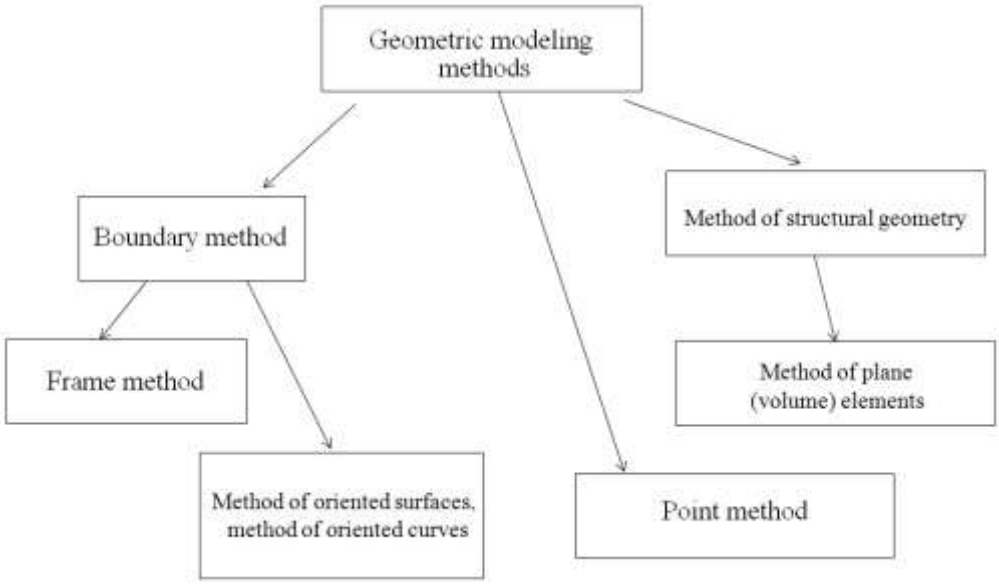


Fig.1. Classification of geometric modeling approaches

A feature of the developed system of geometric modeling, focused on solving the problems of automation of design of engineering objects, is the use of a combination of two approaches to geometric modeling, i.e. the creation of a combined model of the MKOMB object

$$MKOMB=\{MK, BM, F\}, \quad (2)$$

where MKG is an object model using the structural geometry method;
BM – model of the project using the marginal approach;
F – operators convert MKV to BM and BM in MKG.

Geometric models have a hierarchical structure that results from the bottom-up construction. The element of the geometric structure [5, 6] can be: a coordinate system, a point, a curve (flat, spatial), a surface, a geometric body.

A generalized scheme of options for building models of objects is shown in Fig. 2. Arrows indicate how to form geometric elements by using elements of a different dimension.

In the two-dimensional case, the objects of modeling are planar contours consisting of segments of planar curves. The elements of the topological structure of the contours are node points, line segments. The curve can be represented by a set of points, provided that they are close enough to each other.

Point method of geometric modeling uses the model

$$MP=\{P,SA,\}, \quad (3)$$

where P – set of points;
SA – set of attributes $<x_i, y_i.>$, where x_i, y_i – the coordinates of the point;
DIP – display connecting a pair of incidental points.

The MP model is the simplest of the models. It does not use a mathematical description of the curve. The use of mathematical description of the curve complicates the model, but has a number of advantages. In particular:

* mathematical description is accurate, which makes it easy to calculate curve characteristics such as slope, radius of curvature, etc.;

* there is no need for an interpolation scheme to find intermediate points;

* the use of the MP model is difficult if the shape of the curve has to be changed to meet some given criterion.

Mathematically, the curve can be presented in parametric or nonparametric form. The parametric form of the curve description has the form

$$x = f(t),\ y = g(t) \tag{4}$$

where t is the parameter.

A nonparametric curve is described as an explicit or implicit function. The explicit description is $y = f(x)$, implicit $f(x, y) = 0$. A point on an implicit curvilinear segment can be determined by computing the root of an algebraic equation.

An oriented curve is a curve with a specified bypass direction. The orientation of the curve $f(x, y) = 0$ is determined by the direction of the path relative to the area specified by the inequalities $f(x, y) > 0$. Obviously, the curve $-f(x, y) = 0$ coincides with the curve $f(x, y) = 0$, but has the opposite orientation.

The method of oriented curves is to build a geometric model

$$MOC=\{OC, N, SA, DIP\}, \tag{5}$$

where OC-oriented curves;
SA – set of attributes $<x_i, y_i.>$, where x_i, y_i – the coordinates of the node points;
DIP – display connecting a pair of incidental node points.

The approach of structural geometry on the plane is to construct a two-dimensional object (contour) from a set of specified two-dimensional elementary (base) objects by means of geometric operations of Union, intersection and subtraction taking into account the orientation.

As the basic elements can be used: point, line segment, circle arc, arc parabola, arc hyperbola, arc ellipse, rectangle, parallelogram, circle, etc.

The method of plane elements is to build a geometric model

$$MPE=\{EO, CAP, PAR, PO, GO\}, \tag{6}$$

where EO – set of elementary objects;
CAP – the coordinates of the anchor point; <x, y, φ> – coordinates and angles that define the position of the object in the main coordinate system;
PAR – object parameter;
PO – the transformation of the object on the plane;
GO – geometric union and intersection operations.

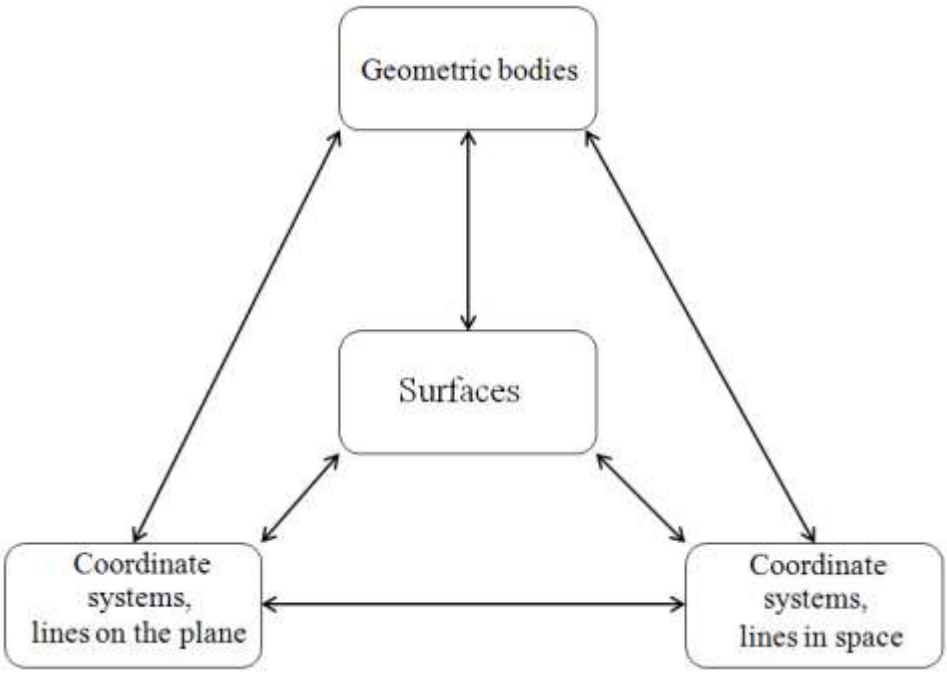


Fig.2. Generalized scheme of options for building object models

Various transformations can be performed on geometric models of objects: transfer, rotation, uniform and non-uniform scaling. Geometric transformations are defined as the objective mapping of a coordinate space into itself. The structure of the object is preserved.

The transformation of objects can be carried out as a transformation of a set of points characterizing the object into another set, provided that these sets are described in the same coordinate system, or as a replacement of the local coordinate system of the object to another coordinate system, which allows to synthesize a set of objects, each of which is described in the local coordinate system, and

The mathematical basis of transformations used in the geometric modeling of objects is the operation of multiplication of matrices that determine the elements of the object and the types of transformation. Transformations of transfer, rotation and scaling in matrix form are written in the form of

$$P' = P + T, \quad P' = PR, \quad P' = PS, \tag{7}$$

where P' is the transformed coordinates of a geometric element;
P - transformable coordinates;
T, R, S – respectively matrix transfer, rotation and scaling.

When modeling objects can be used combinations of the basic types of geometric transformations (translation, rotation, scaling) and various derivatives thereof: mirroring, copying, etc.

Design objects can be modeled interactively in a machine graphics system using a problem-oriented language or a software package. In accordance with the chosen method of modeling modification of objects is carried out by technical devices of the system with the use of embedded software and hardware, language operators and programs of the package.

Modification of objects is carried out with their linguistic models, with internal forms of representation of object models, with images.

Analysis of geometric models of design objects allows to determine their size, area, volume, centers of gravity, moments of inertia, static moments, distances between objects, etc.

One of the main areas of application of standard methods of geometric modeling – computer-aided design and automated control systems of technological processes of design and production of mechanical and Electromechanical systems.

The boundary method of object modeling is appropriate to use in areas where dynamic surfaces are designed, i.e. surfaces interacting with the external environment, or surfaces to which high aesthetic requirements are imposed. Dynamic surfaces are divided into two classes:

* described by the environment
  (external contours of aircraft, ships, cars, turbine blades, etc.));
* tracer-guiding medium
  (air and hydraulic channels, spiral chambers and turbine suction pipes, etc.).

Methods of structural geometry are used in systems focused on the design of machine-building structures, parts obtained in the technology of stamping, cutting, etc. This type of model is most adequate to such processes of designing geometric shapes of parts, when the designer forms a part as a combination of basic elements of shape (cylinder, cone, prism, etc.) or collects it from some functional elements, for example, the shaft section for planting a gear, etc.

For low-performance systems, a method that uses wireframe models is used. Frame models are usually used to define objects representing polyhedra, i.e. closed polyhedra of arbitrary shape bounded by plane faces. Wireframe representation is often used when displaying objects as one of the visualization methods. Feature images of objects represented by a point frame, linear discrete frame and mesh frame.

The efficiency of the use of geometric models of objects in computer-aided design depends on the means of data storage and processing. The composition of the database is determined by taking into account the characteristics of the process and objects of design and current standards and reference data.

Graphic displays are widely used for visual inspection and modification of designed objects. Two types of graphic displays can be distinguished by scanning technology: vector and raster. When displaying object models on a vector display, it is sufficient to use a wire representation; visualization of patterns on a raster display allows you to get the color "deep" images of objects.

Drawing documentation is one of the main results of computer-aided design systems. The complexity of its formation depends on the completeness of the data in the geometric model. For example, a dimensional grid saved during the simulation process allows you to automatically place dimensions in the drawing.

To design a drawing of a part represented as a model of a three-dimensional geometric object, the following tasks are required: to determine the number of projections required to represent the part; to highlight the main view; to build the appropriate projections; to determine the scale of the image in the drawing field; to place dimensions and technological information; to apply technical requirements; to fill in the drawing stamp.

The most common form of application of automated systems of technological preparation of production is programming for machines with numerical control. Based on the generated object model, programs for CNC machines can be created much faster as the object geometry is already defined. In the automated systems of preparation of control information for CNC machines the most convenient way to use a geometric model, formed by the method of oriented surfaces.

2 Geometric modeling examples

De Casteljo algorithm was chosen for the implementation of the planned program for the design of stretch ceilings. The program is designed "Euroceiling" is a window size of 1000 by 600 pixels.

The menu is located at the top:

- Order information
- Drawing
- Print
- Settings
- Help

At the bottom there are quick start buttons that allow you to quickly build a particular shape with the specified parameters.

Also, on the same panel there are buttons "Turn" and "Step by step". The first button allows you to rotate the selected object. The second button allows you to turn on the step-by-step construction mode in order to carefully consider the geometric modeling process and adjust it if necessary.

On the left is the information coordinate panel, where the coordinates of all edges of the drawing are displayed in a table view.

Let's consider the construction based on diagonals. If the construction is done by entering the sides and diagonals, then you must specify the number of angles and press the "Build" button, then a table and a tree of sides will be formed to enter the length of the values in centimeters (figure 3).

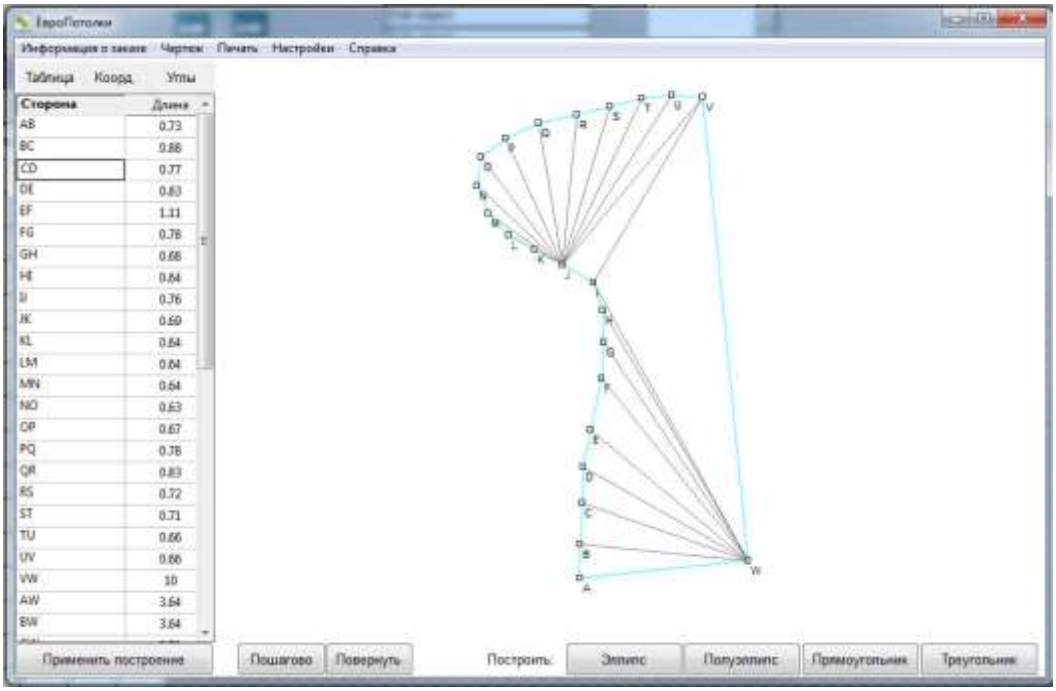


Fig.3. Drawing Based on the sides and diagonals.

In this case, the edges of the constructions are denoted by capital Latin letters, for example, A B (the outer edge between points A and B), AW (the inner edge, which is used to "bind" pieces of the ceiling together).

There are three ways of specifying lengths of sides and diagonals:

Table - the names of the sides are put down automatically, but the name of the diagonals have to be entered manually;

Tree-contains all variants of sides and diagonals, which avoids manual input of the sides and diagonals, also provides automatic transition to the next size in order when you enter the current one and press "Enter" (this applies to both sides and diagonals);

Angles-avoids entering diagonal lengths if known angles (in degrees) between points, for example, there is an ABCD rectangle (AB=300;BC=500;CD=300;DA=500) and it is known that the angle A=90 degrees, hence the diagonal BD=583.1, just note that entering an angle in degrees only automatically calculates the corresponding diagonal.

If you enter an excessive number of diagonals and there are differences in size, an error can be generated, which is displayed on the screen.

The program has a number of support functions:

* Function to "Rotate" - rotates the drawing 90 degrees clockwise;
* The "Build step" - allows you to build a drawing step by step "point by point", thus in difficult situations to assign the responsibility of drawing on the user.

After drawing, you can manually adjust the desired point in the drawing, it is necessary to hold the left mouse button on the changing point and not releasing the move to the desired location. After manual correction of the drawing, an error formed during the construction of the drawing is formed or recalculated.

Consider the construction based on an ellipse. If the construction will be based on an ellipse or circle, you must specify the number of angles and two radii (for the circle radii will be the same and the second radius can not be specified to leave equal to 0), then you must click "Build Ellipse" or "Build half-Ellipse" (Figure 4). The program will automatically build the drawing based on an ellipse.

It is possible to build a half-ellipse, you also need to specify two radii, the number of vertices and click "Build half-Ellipse" (figure 5). If you need to build the right circle, it is enough to specify two of the same radius (figure 6).

Consider manual construction of the drawing. To work with designer ceilings, which are already drawn in special modeling programs with the preservation of all proportions, you can use the "Manual construction" (figure 7). It is possible to enter the data (name of a line and its length) directly in the table. To apply the function, click the «Apply build» button.

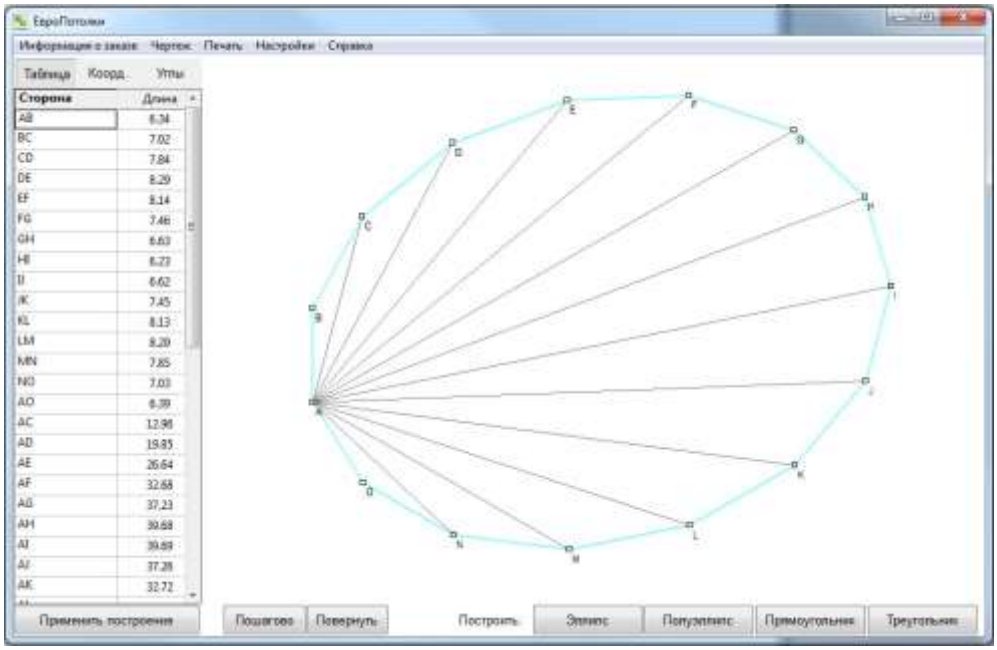


Fig.4. Building an ellipse

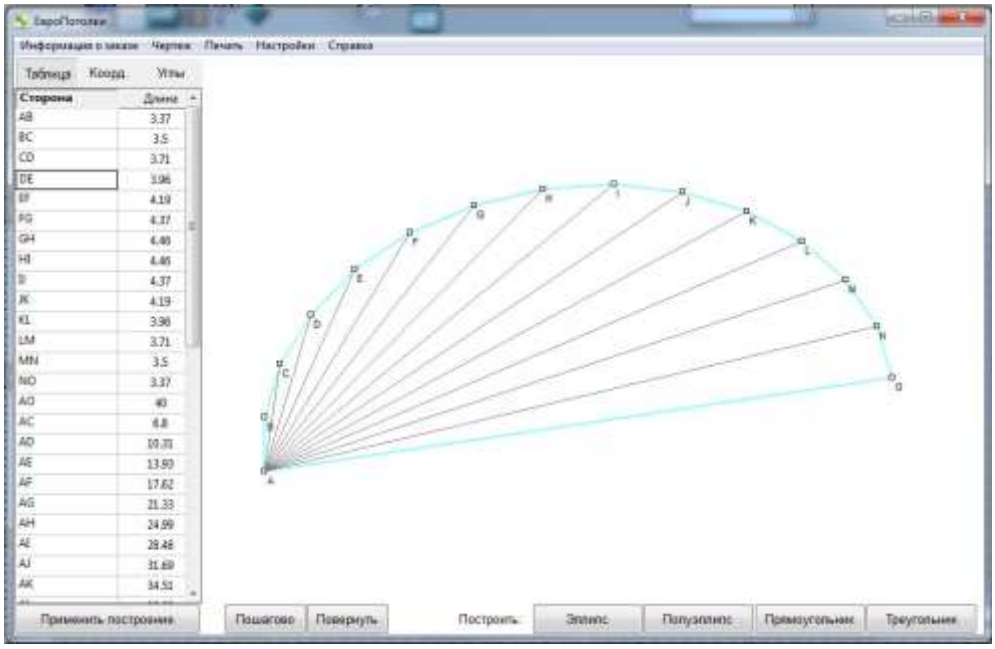


Fig.5. Construction of the half-ellipse

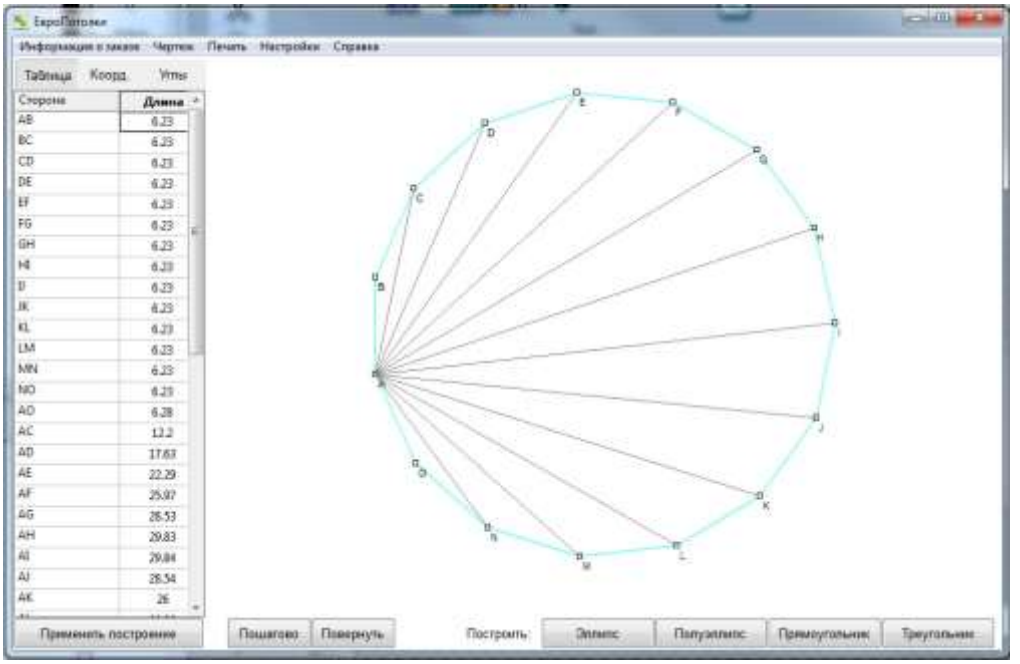


Fig.6. A perfect circle

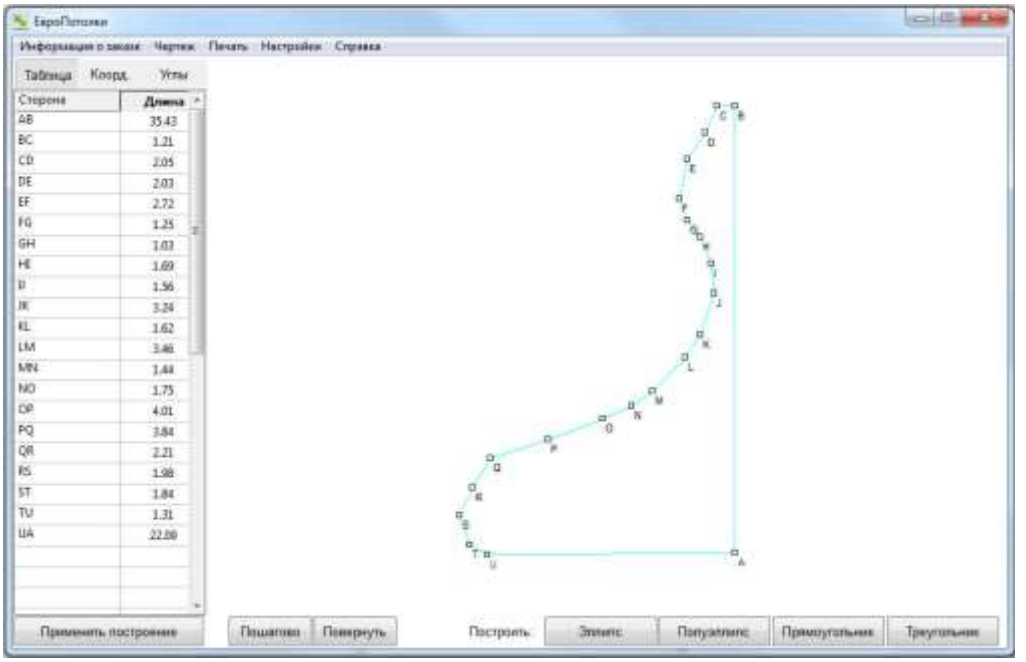


Fig.7. Manual construction mode

Conclusion

As a result, a system of designing stretch ceilings was developed, which has the following features: Perform a construction based on the input data table with coordinates and angles; Execution of construction of an ellipse; Performing a half-ellipse build; The build of the rectangle; The build triangle; Rotation functions of selected edges; Step-by-step build function; roject print function;

References

[1] Patrikalakis N. M. Surface-to-surface intersections // *IEEE Computer Graphics and Applications*, 1993, 13(1):89–95.

[2] Hirsch M.W. Differential topology. Graduate texts in mathematics. Springer-Verlag, Berlin, 1976.

[3] Chavel I. *Eigenvalues in Riemannian Geometry. Pure and applied mathematics*. Academic Press Inc., Orlando, 1984.

[4] Zyablov E.L., Preobrazhensky Yu. P.Building an object-semantic model of the control system // *Bulletin of the Voronezh Institute of high technologies*. 2008. No. 3. P. 029-030.

[5] Panevin R.Y., Preobrazhensky Y. P. Structural and functional requirements for software complex knowledge representation // *Bulletin of the Voronezh Institute of high technologies*. 2008. No. 3. P. 061-064.

[6] Chutchenko Yu. E., Preobrazhensky A. P. Study the possibility of improving the image quality // *The territory of science*. 2007. No. 3. P. 364-369.

Prof. Yakov Lvovich
Doctor of Sciences (Engineering), professor,
Voronezh Institute of High Technologies, Russia
office@vivt.ru

Assoc. Prof. Alexander Plotnikov
Candidate of Sciences (Engineering), associate professor,
Voronezh Institute of High Technologies, Russia
office@vivt.ru

Ing. Juraj Štefanovič, PhD.
Faculty of Informatics, Pan-European University, Bratislava, Slovakia
juraj.stefanovic@paneurouni.com

DEVELOPMENT OF TECHNIQUE FOR RECOGNITION OF THE COMPLEX SHAPE SIGNAL IMAGES

Oleg Choporov, Andrey Preobrazhenskiy, Eugen Ružický

Abstract:

With the help of digital image conversion, they create conditions for improving the perception of the image, the formation of a certain artistic image, the allocation of informative features. Image conversion is carried out by various methods, including optical, photochemical and digital methods. The latter are becoming increasingly common, as technical capabilities only grow from year to year. The purpose and objectives of this work: to analyze the main methods of image processing and recognition; to develop an effective algorithm for recognizing signal images; to develop a software product that conducts experiments to recognize complex waveforms. Information about the degree of similarity of the two signals is shown at their maximum coincidence and is determined by the behavior of the correlation integral (functional) at the extreme point, that is, the value of the integral and its derivatives. This functional is an energy characteristic, the value of which is determined by the brightness and area of the image visible through the standard. Its extremum is observed in cases where the standard is fully fit in the image, and the coincidence of their shape is not necessary. Therefore, this value characterizes the similarity of the image and the standard in form ambiguously. The absolute value of the second derivative of the correlation integral to the point of its extremum is preferable for estimating the similarity of the image and the standard in form. The dependence of the functionals on the degree of "blur"is presented. A necessary condition for the practical value of such recognition algorithm is the independence or weak dependence of the functional threshold value on the shape, area and contrast of the image. The table shows the values for some images with different area and contrast. as an indicator of the similarity of the image and the standard in form, it is advisable to use the module of the second derivative of the correlation integral normalized in contrast at the extreme point. First, the program receives the original image in bmp or jpeg format, and shows it to the user. The image is then muted according to the noise type selected by the user and the parameters set for the selected noise type, then displayed on the form. You can save an image, run an experiment (plot), define a shape, or noise the image again. To compare the capabilities of the algorithm under study and the human visual system according to the classification of "blurred" images, the images of simple geometric shapes are shown recognized by an automatic device implemented in the form of a software tool. The type of distortion that can be used: monochrome noise by Gauss, color noise by Gauss, spray, turn into mosaic, blur. The program compares the distorted image with each of the available reference figures, estimates the total standard deviation, and concludes what image is depicted, based on its minimum value.

Keywords:

Image cognition, algorithm, program

ACM Computing Classification System:

Computer vision, Human computer interaction, Information systems applications

Introduction

Currently, digital image conversion is widely used in measurement and applied television systems and broadcast television. With the help of digital image conversion, conditions are created to improve the perception of the image, the formation of a certain artistic image, the allocation of informative features (for example, in object recognition systems), etc. Conversion is carried out by various methods, which include optical, photochemical and digital methods [1, 2]. The latter are becoming increasingly common, as technical capabilities only grow from year to year. Despite the fact that other methods are a little passed the position, all these are of independent interest in communications technology, computer technology, medicine, the environment, the conversion of film images and other various fields.

A number of approaches to the study of images is based not only on the use of a priori data about the optical characteristics of the image, but also on the possibilities of implementing the analysis procedures. For example: the use of image analysis procedures to control chip patterns, in medicine and biology in the analysis of blood smears, etc.

Purpose and objectives of this work:

- To analyze the main methods of image processing and recognition;
- Develop an efficient algorithm for signal image recognition
- To develop a software product that conducts experiments to recognize complex waveforms.

1 Image recognition algorithm

The initial data for the classification of images of structural or physical features are the results of the measurement of the latter. The most informative stable characteristic of a simple image is its shape, as the color and brightness distribution of the image in real conditions are variable and are not always class features. Under the form in General, the external outline, external view, contours of the object. For a two-dimensional image, it is difficult to define the shape as a specific trajectory for each class of images in the form of a closed line in the relative coordinate system. Unlike video and radio signals, the form of which determines the noise immunity of reception, measurement accuracy, resolution and other frequency-energy factors, the form of images is essentially a visa-al code, which contains information about the name, purpose and other information about the object of observation.

Closed lines can have an infinite number of trajectories, which creates a significant redundancy in this encoding of information. Perhaps that is why, when perceiving contour images, the visual analyzer of the human brain easily completes the missing fragments of the form and perceives the set of contours as a complete image [3, 4].

The peculiarity of the classification feature is the impossibility of its evaluation, since the signal code, including the shape of the image, is not subject to measurement. To evaluate only the degree of conformity (similarity) with the same signal structure (the benchmark).

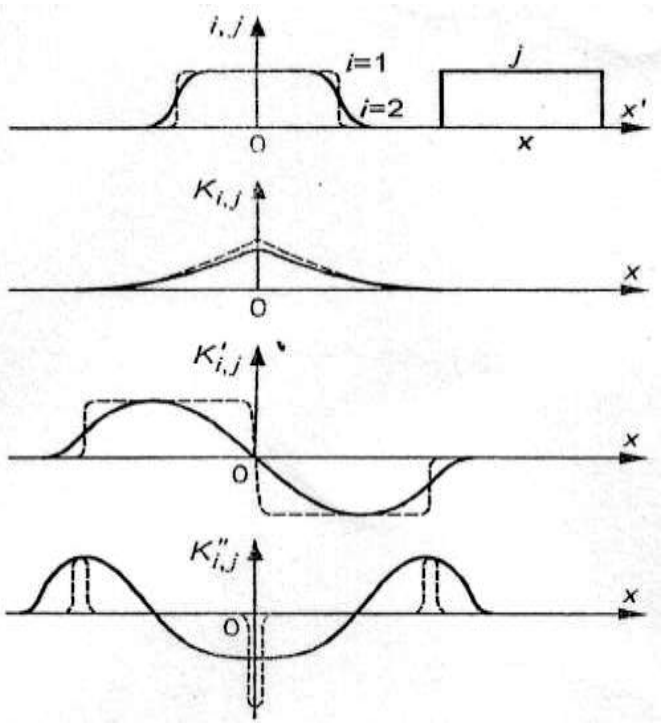


Fig.1. Definition of a parameter containing information on the similarity of the signal and the standard in form

Fig.2. Geometric figures presented for recognition

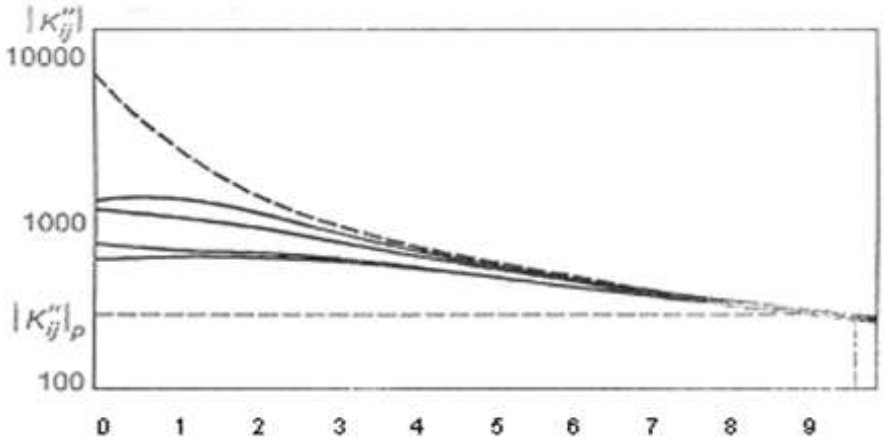


Fig.3. Dependence of the trapezoid similarity with the standards on the degree of its "Blur»

Presumably, the information about the degree of similarity between the two signals i(x, y) and j(x, y) appears at their maximum coincidence and is determined by the behavior of the correlation integral at the extreme point, i.e. the value of the integral (KI) and its derivatives (1)

$$Kij(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} i(x',y')\,j(x'-x,y'-y)dx'dy' \qquad (1)$$

To identify the parameter containing the required information, we analyze the dependence of these characteristics of the correlation integral on the distortion (change) of the image shape. To simplify the analysis, let's consider a one-dimensional signal passing through a low-pass filter.

Figure 1 shows two distorted rectangular pulses of different degrees i_1, j_2, the reference pulse *j*, their correlation integral K_{ij}, its first K'_{ij} and second K''_{ij} derivatives. It follows from figure 1 that the odd derivatives at the extreme point of the correlation integral are zero, and the even ones depend monotonically on the degree of "blur" of the momentum. Therefore, for comparative analysis it is advisable to allocate the values of the correlation integral and its second derivative at the extreme point.

Functional K_{ij}, it is an energy characteristic, the value of which is determined by the brightness and area visible through the standard image. Its extremum is observed in cases where the standard is fully fit in the image, and the coincidence of their shape is not necessary. Therefore, the value K_{ij} characterizes ambiguously the similarity of the image and the standard in form.

The second derivative of the correlation integral $K''ij$ essentially depends on the correspondence of the signal form, striving for the absolute value to infinity with perfect similarity (see Fig. 1).In this case, the change in the image area can not be compensated for its non-compliance with the standard form, since the value $\left|K''ij\right|$ decreases due to the formation of a flat portion of the top of the correlation integral.

Therefore, the absolute value of the second derivative of the correlation integral to the point of its extremum is preferable for estimating the similarity of the image and the standard in form.

We will analyze the peculiarities of using the functional $\left|K''ij\right|$ in recognition problems by the example of classification of the geometric shapes depicted in figure 2. Image ν of each figure with the help of a graphic editor Adobe Photoshop formed the standard of the file in the form of a file in the Windows BitMap format. In the course of research, the images were gradually distorted using the developed program (see below) and the similarity of the obtained images i with each of the standards j (2) was evaluated.

$$i(x,y) = \frac{1}{2\pi\delta^2}\int_{-\infty}^{\infty} v(x_1,y_1)e^{-0,5\frac{(x-x_1)^2+(y-y_1)^2}{\delta^2}}dx_1dy_1 \qquad (2)$$

In Fig. 3 we can see the dependence of the functionals $\left|K''ij\right|$ on the degree of "blurring" of the trapezoid is characterized by parameter σ. From figure 3 it can be seen that the similarity of a trapezoid with its own benchmark (dotted line) is dominant to the values of $\sigma \approx 9{,}4$ the pixels in which the dotted line and one or more solid lines, describing an image similar to nonproprietary standards intersect.

The corresponding intersection point of the functional value will be called the threshold ($\left|K''ij\right|_p$). Further "blurring" of the image entails a decrease in its similarity with its own pattern in comparison with some of the improper patterns.

As shown in the figure. 3 the dependences are typical for all the geometric shapes under study, the classification of the presented image involves an assessment of its similarity with each of the standards, the choice of the largest functional $\left|K''ij\right|$ and its comparison with a non-threshold. If the value of $\left|K''ij\right|_{\max}$ is greater than the threshold shown, the image of the standard corresponding to this functional is considered, if it is equal to or less than the threshold, the recognition is rejected.

It is easy to show that the necessary condition for the practical value of such recognition algorithm is the independence or weak dependence of the functional threshold value $\left|K''ij\right|$ on the shape, area and contrast of the image. Table 1 shows the values $\left|K''ij\right|_p$ for some images with different areas of S and contrast, expressed in pixels and machine units, respectively. According to these data, the threshold $\left|K''ij\right|_p$ is slightly dependent on the shape and area of the image [5, 6], but is proportional to the absolute value of its contrast. The normalization of the functional $\left|K''ij\right|$ on image contrast allows to eliminate this dependence, as evidenced by the data in table 2.

Table 1-Functional thresholds $\left|K''ij\right|$

Image	S	ΔB		
		50	-50	200
Square	625	99.5	99.5	398.1
	2500	108.	108.	435
	10000	112.6	112.6	450.6
Right triangle	625	99.9	99.9	399.8
	2500	98.7	98.7	394.9
	10000	96.3	96.3	385.2
Rhombus	625	92.7	92.7	370.7
	2500	111.3	111.3	445
	10000	110	110	439.9

Table 2-Threshold values of the contrast normalized functional image $\left|K''ij\right|$

Image	S	ΔB		
		50	-50	200
Square	625	2.82	2.82	2.82
	2500	2.89	2.89	2.89
	10000	2.88	2.88	2.88
Right triangle	625	2.81	2.81	2.81
	2500	2.79	2.79	2.79
	10000	2.78	2.78	2.78
Rhombus	625	2.78	2.78	2.78
	2500	2.87	2.87	2.87
	10000	2.88	2.88	2.88

Thus, as an indicator of the similarity of the image and the standard in form, it is advisable to use the module of the second derivative of the correlation integral normalized in contrast at the extreme point.

For the convenience of further presentation, we denote: W_{ij} - the similarity index of the image i and the standard j; W_v -the threshold value of the indicator W_{ij} at $i = f(v), j \in J$; J the set (bank) of patterns; W — the threshold for deciding on the compliance of the presented image i to a particular standard. With this in mind, the decision rule can be written as follows (3), where q is the image of the standard, in which the similarity index W_{ij} reaches its maximum value on the set of standards J.

$$\begin{cases} q = \arg\max(W_{ij}) \\ \quad j \in J \\ \quad v = q \\ \quad Wiq \le W, \\ \quad v \notin J \end{cases} \tag{3}$$

The reliability of image recognition is determined by the value of the threshold W. to avoid incorrect classification, the threshold W must be equal to or greater than the maximum for the set of images V threshold W_v, where V is the set of images of objects of observation corresponding to the set of standards J.

$$W_{max} \ge_{v \notin V}^{max} (W_v) \tag{4}$$

In this case, the capabilities of the algorithm to recognize distorted images are somewhat reduced. Indeed, the majority of images v individual threshold recognition W_v will be less than the total threshold W, but when $W > Wig > W_v$ will be a refusal of recognition, although the image v can still be classified correctly.

When the W threshold is reduced, there is a possibility of incorrect recognition of images for which $W_v > W$,, but the degree of distortion at which these images can be correctly recognized increases. This condition (4) applies to all images [7, 8] if the W threshold is less than the minimum individual W_v threshold. If recognition rejection is not allowed, i.e. the image must be classified at any distortion degree, the W threshold is equal to zero.

First, the program receives the original image in bmp or jpeg format, and shows it to the user. The image is then muted according to the noise type selected by the user and the parameters set for the selected noise type, then displayed on the form. You can save an image, run an experiment (plot), define a shape, or noise the image again.

If you choose to save an image, the first step is to check whether the image can be saved, then the image is saved to a file with the specified name and extension.

If an experiment is selected, the number of tests, the type of noise, the noise parameters are determined, then the increment to the noise degree parameter is determined, and for each of the reference figures a series of experiments is carried out:

* The image is noisy from the minimum to the specified noise level
* After each noise is estimated KI and stored in an array of values KI.

After that, the values of KI are displayed on the chart, where each line corresponds to the KI values at different degrees of image noise.

If you select image definition, then for each reference figure is determined to the original relative to the reference, then select the minimum value of KI and remembered for which figure this value was the minimum. When this is done, the program will show a sample of the reference shape and display the shape name on the screen. You can also save the resulting image.

If you select the re-noise, the noise is carried out relative to the already noisy image, otherwise the noise is carried out for the original.

2 Recognition results

To compare the capabilities of the algorithm under study and the human visual system according to the classification of "blurred" images, figure 4 and 5 show images of simple geometric shapes (area S = 625 pixels) recognized by an automatic device implemented as a software tool. For the images in figure 4, the recognition threshold $W = W_{max}$, in figure 5, the images were distorted to their individual W_v thresholds. It is obvious that the images in figure 4 are shown to a lesser extent than in figure 5, which is a fee for excluding their incorrect recognition at $W = W_{max}$. In General, according to figure 4 and 5, it can be concluded that within the framework of this task, the efficiency of automatic recognition of blurred images is close to the efficiency of the human visual system

Thus, the above-proposed measured index of similarity between the image and the standard in form provides a sufficiently high efficiency of recognition of distorted images when passing through optical systems or media without taking into account a priori information about the degree of their "blur".

Fig.4. Images distorted to individual thresholds W_v:
1 - square; 2 - circle; 3 - diamond; 4 - trapezoid; 5 - right triangle.

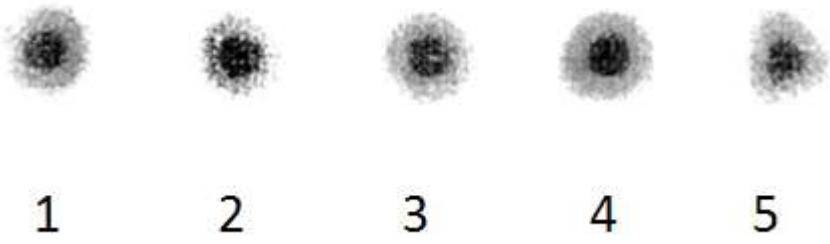


Fig.5. Images distorted to individual thresholds W_v:
1 - square; 2 - circle; 3 - diamond; 4 - trapezoid; 5 - right triangle.

For each type of noise, there is a certain value of the noise level at which the program stops correctly determining the reference images For noise type "Gaussian monochrome" the following threshold values were obtained:

Square-degree 4200; KI value-1,1300512728883
Round - the degree of 4130; a value of KI - 1,12543698834234
Rhombus - grade 4400; the value of KI - 1,12643149912134
A-line - degree 4350; a value of KI - 1,13357093373766
Right triangle - a 4100 degree; the value of KI - 1,126781459449

«Gaussian color»:

The square - degree of the 4400; the value of KI - 0,91648247891954
Round - the degree of 4350; a value of KI - 0,918156459591398
Rhombus-degree 4900; value KI-0.917142759509086
A-line - degree 4300; the value of KI - 0,915388786370116
Right triangle - degree of 4050; a value of KI - 0,915354674331766

«Gaussian blur»

Square-degree 19; KI value-0,394154110055535
Round - the degree of 21.5; a value of KI - 0,408772649106782
Rhombus-degree 22; value KI-0,372755526395548
A-line - degree 14; the value of KI - 0,361606294297676
Right triangle-degree 24; value KI-0,402003509124454

«Spray»

Square - grade 3; the value of KI - 0,601763233722686
Round - the degree 5; the value of KI - 0,668544297411066
Rhombus-degree of 8; value KI-0,646404502806758
A-line - degree 4; the value of KI - 0,669168646834291
Right triangle - the degree 6; the value of KI - 0,606404784725536

«Mosaic»

Square-degree 20; KI value-0,273543905599172
Round - 30 degree; the value of KI - 0,378621893443313
Rhombus - degree 40; the value of KI - 0,40252055209977
A-line - degree 30; the value of KI - 0,388925655293391
A right triangle is 35 degree; the value of KI - 0,317185860370666

3 Software implementation of the algorithm

When you start the application, at first you need to open the image (Fig. 6).

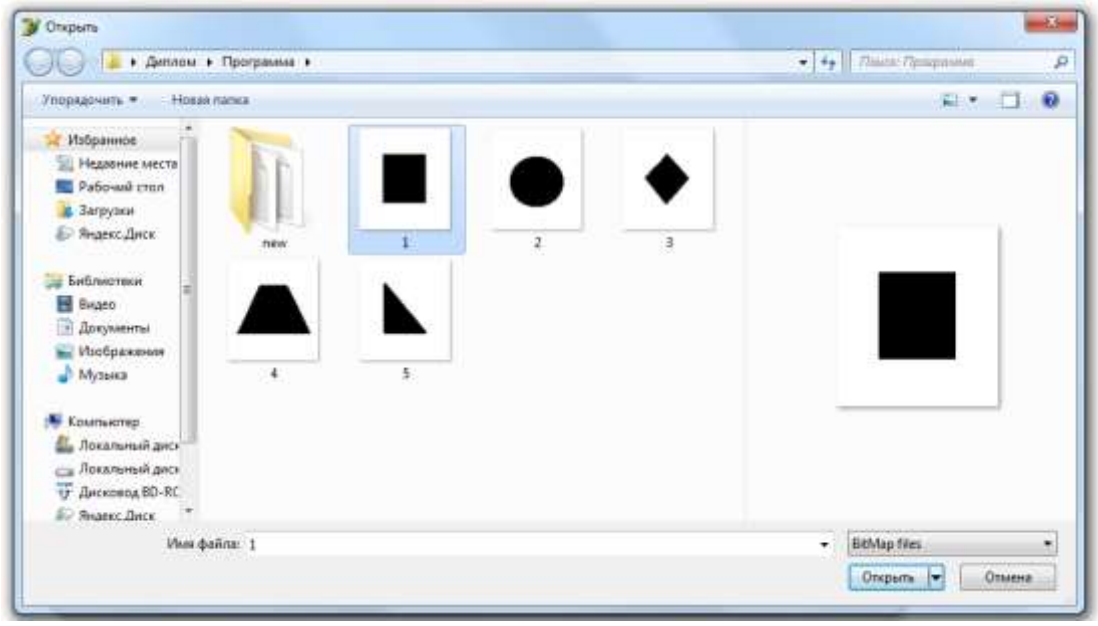


Fig.6. Opening the image

When you select an image, you can specify which images to open, bmp or jpeg. Thus, we exclude the possibility of opening files with the wrong format. After opening the image, the contents of the form will change as follows (Fig. 7).

The selected image will be displayed on the left side of the form, and information messages about what will be displayed on them will appear in the middle and right panels of images. After opening the image you need to make the original image distortion, for further definition of the image. To do this, it is proposed to choose the type of distortion that will be produced:

* Monochrome noise over Gauss
* Color noise by Gauss
* Spray
* To turn into a mosaic
* Blur

For the convenience of users, the primary parameters for all noise algorithms have already been made, but the user can change them at any time. Choosing the right panel the type of noise, at the top right of automatically changing the panel settings for the type of noise.

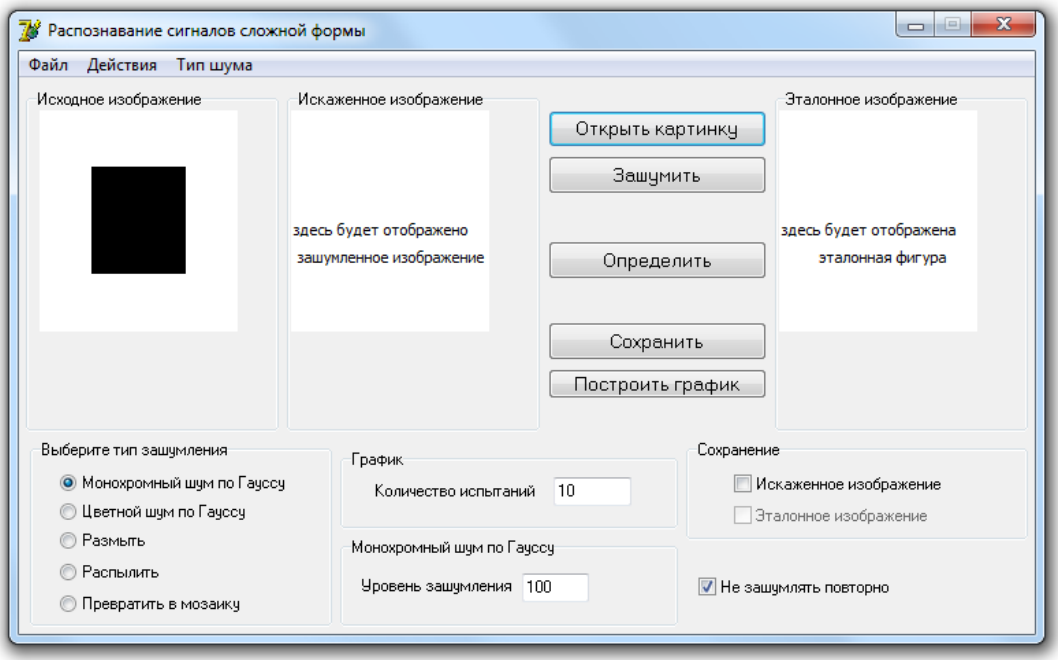


Fig.7. Adding a reference standard

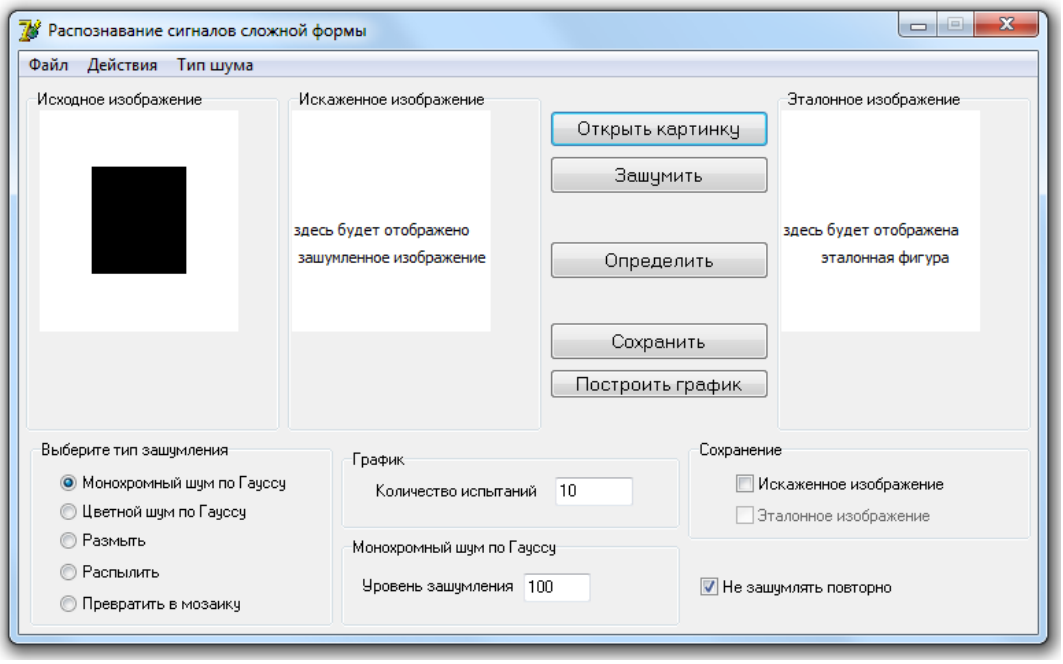


Fig.8. Variant for Gaussian monochrome noise.

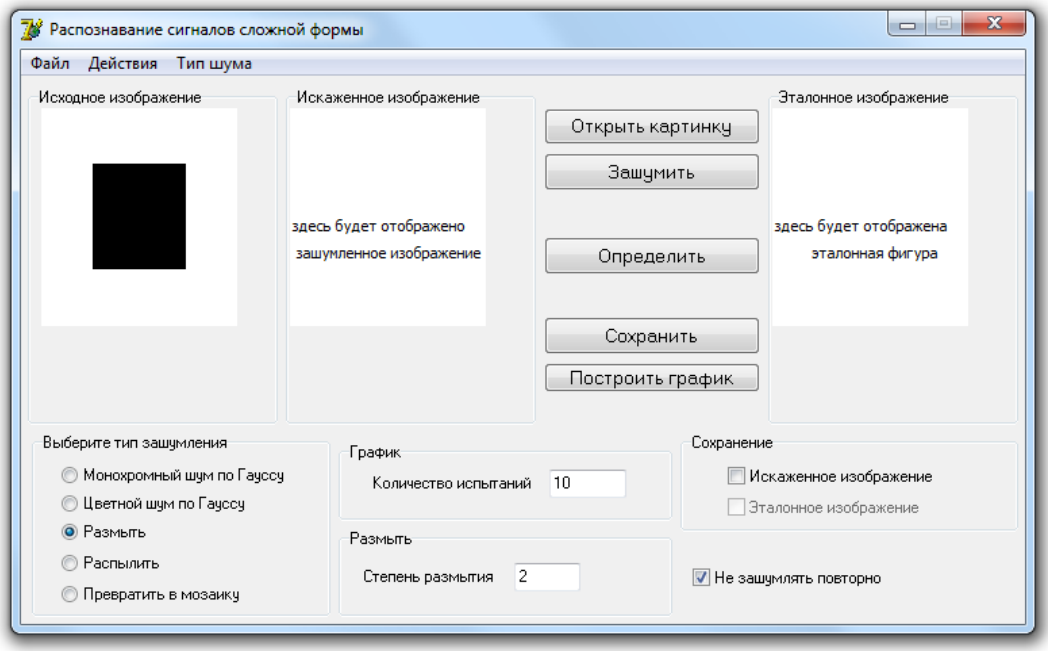
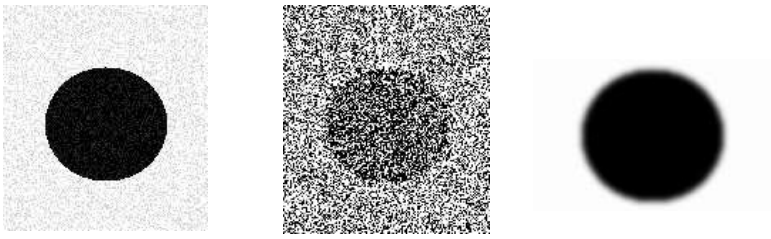


Fig.9. Changed the parameters panel for "Blur" noise.

Thus, we can adjust the noise level for each type of noise, for example (Fig. 8), (Fig. 9).

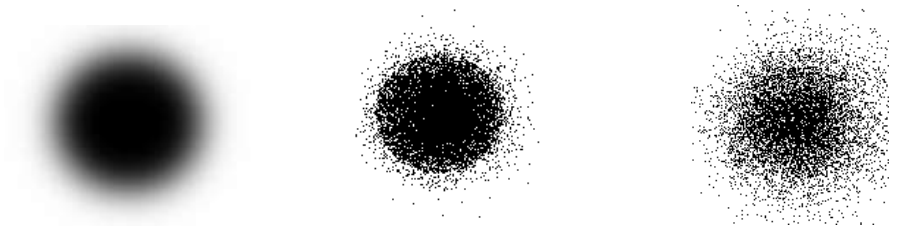
If the "Monochrome noise by Gauss" is considered, the points of white or black color are randomly applied to the image (Fig. 10 (a). By changing the noise level, we get a more or less noisy image (Fig. 10 (b). Image blur may be present (Fig. 10 (c).



a) b) c)

Fig.10.
a) Monochrome noise over Gauss.
b) Monochrome noise over Gauss, with stronger noise.
c) Gaussian blur.

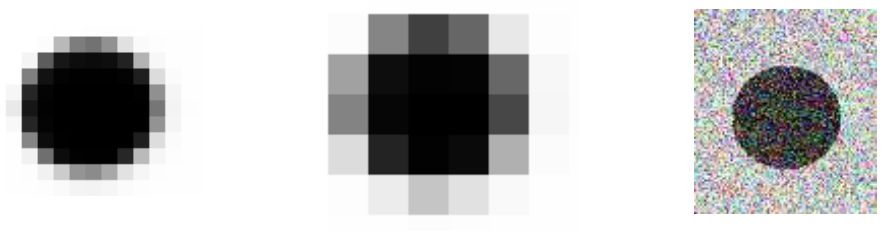
When increasing the degree of blur we get a more blurred image (Fig. 11 (a)). Analog of the "spray" effect in MS Paint (Fig. 11 (b)). Increase the depth of spraying (Fig. 11 (c)).



a) b) c)

Fig.11.
a) Gaussian blur, with a stronger degree of blur.
b) «Spray» Effect.
c) «Spray» effect, with increasing depth.

If you consider a "mosaic", the image is divided into rectangles of a given size and painted in medium color, using the arithmetic mean of the components. (Fig. 12).



a) b) c)

Fig.12.
a) Mosaic
b) Mosaic, when increasing the size of the rectangles
c) Gaussian noise with "color" noise

Thus, we see that by choosing the type of noise and adjusting the noise parameters, we can get different degrees of image distortion, which can continue to work.

It may be of interest to combine different types of noise, i.e. noise sequentially several types of noise.

To do this, you need to remove the check mark from the item "do not noise again".

4 Definition of a distorted image

Once all the necessary manipulations are carried out, you can proceed directly to the definition of the image. (fig.13).

The principle of estimation of the total standard deviation from the reference figure is used for determination.

The program compares the distorted image to each of the available benchmark figures, estimates of the total standard deviation, and draws a conclusion about what the image depicts, on the basis of the minimum value of standart deviation. The reference figure, when compared with which this minimum standart deviation was obtained, is recognized as the image on the distorted image.

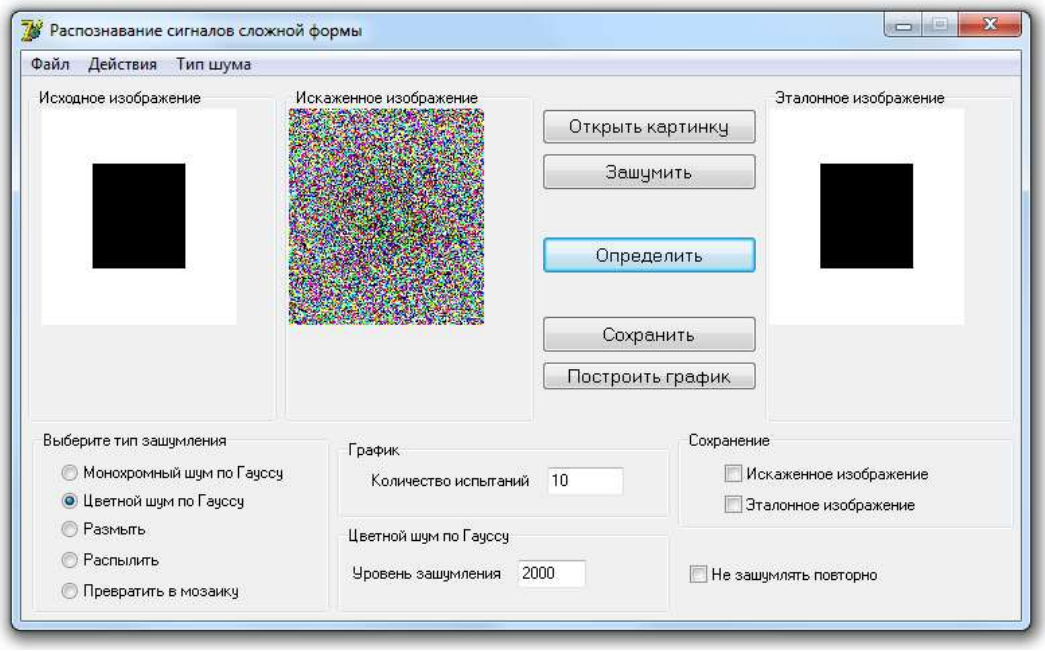


Fig.13.-Image Recognition

The figure that was recognized as the reference for the current distorted image is recognized as the reference for the current image, is displayed in the "Reference image" panel and its name is displayed in the lower part of the main window.

The program provides for the possibility of studying the dependence of the values of standart deviation in relation to different reference figures, depending on the type of noise and noise parameters.

To conduct tests and plotting, you must select the type of noise for which the dependence will be studied, enter noise parameters, and the number of tests.

Tests are conducted in the intervals of noise parameters from the minimum to the specified.

Each line on the graph corresponds to the standart deviation of the distorted image from the reference. (figure 14).

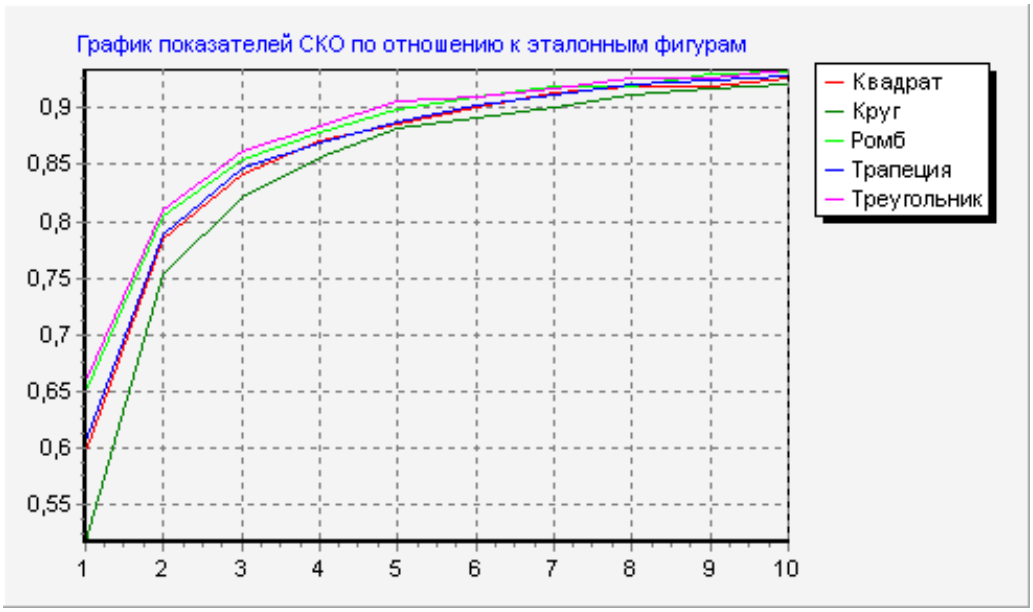


Fig.14. Graph of standart deviation

The graph shows that the more noise occurs, the closer the values of standart deviation for different reference figures. Another example (distortion circle, the noise is "mosaic", when values of the 50x50) (Fig. 15). Each image can be saved with the name selected manually or with the program.

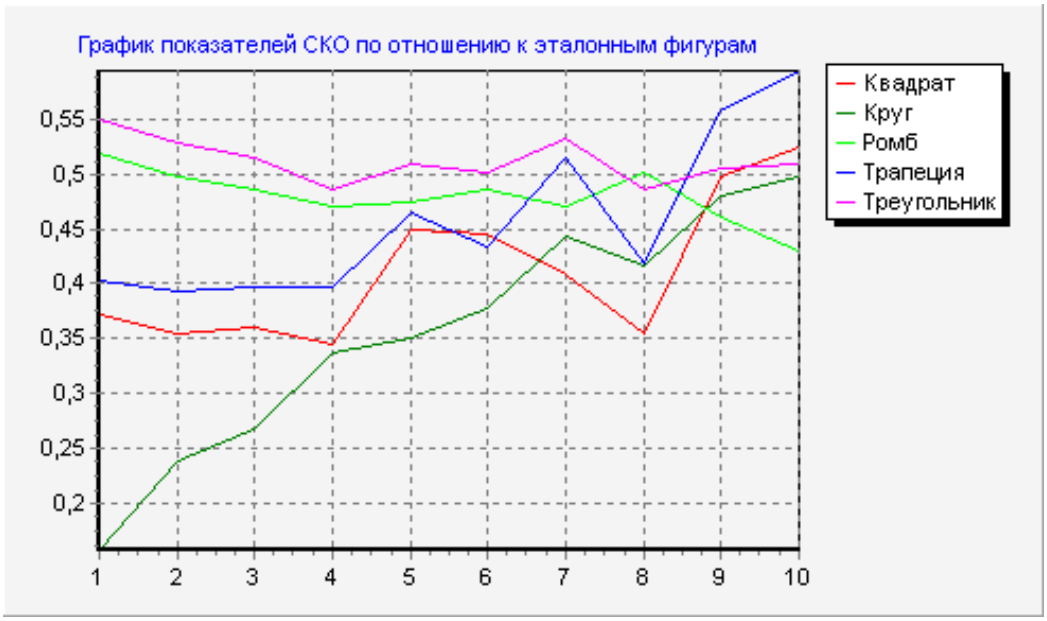


Fig.15. Graph of standart deviation the distortion of the circle.

Conclusion

The model and the algorithm of image recognition based on the correlation approach are constructed. The results of recognition for different types of images and noise are obtained.

On the basis of the developed recognition algorithm, a software product is created in the programming environment. This product allows you to analyze the possibility of recognition for different types of noise and has a fairly simple and intuitive interface.

References

[1] Ermolova V. V., Preobrazhensky N.. Preobrazhensky, Yu. p. System of pattern recognition in clinical practice decision support // Information Technology Applications. 2013. No. 4. P. 55-64.

[2] Chubchenko Yu. E., Preobrazhensky A. P. Study the possibility of improving the image quality // the territory of science. 2007. No. 3. P. 364-369.

[3] Zyablov E. L., Preobrazhensky Yu. P. construction of object-semantic model of the control system. Vestnik Voronezh Institute of high technologies. 2008. No. 3. P. 029-030.

[4] Ty J., Gonzalez R. principles of pattern recognition-M.: Mir, 1978.

[5] Zazulin A.V., Preobrazhensky Yu.P. features of construction of semantic models of the subject area. Vestnik Voronezh Institute of high technologies. 2008. No. 3. P. 026-028.

[6] Panevin, Y. R., Y. P. Transfiguration of the Structural and functional requirements for software complex representation of knowledge // Bulletin of Voronezh Institute of high technologies. 2008. No. 3. P. 061-064.

[7] Dubovik A. S. Applied optics. - Moscow: Nedra, 2002.

[8] Zakaznov N. P. Applied optics. - M.: Engineering, 2002.

Prof. Oleg Choporov
Doctor of Sciences (Engineering), professor,
Voronezh Institute of High Technologies, Russia
choporov_oleg@mail.ru

Prof. Andrey Preobrazhenskiy
Doctor of Sciences (Engineering), associate professor,
Voronezh Institute of High Technologies, Russia
app@vivtl.ru

Assoc. Prof. Eugen Ružický, PhD.
Faculty of Informatics, Pan-European University, Bratislava, Slovakia
eugen.ruzicky@paneurouni.com

AUTOMATED WORKPLACE FOR MANAGING MULTI-LEVEL ACCESS TO INFORMATION AND PROGRAM RESOURCES OF EDUCATIONAL ORGANIZATIONS

Yakov Lvovich, Vera Kostrova, Juraj Štefanovič

Abstract

Educational institutions are increasingly faced with the fact that there are problems of violations of the protection parameters of information and software, they are determined by the influence of unauthorized actions of legal users – employees of the institution.

The need for comprehensive protection of information and software resources determines the development of an integrated system that is integrated into the main information environment of educational institutions, its main focus is related to the prevention of the maximum number of types of "insider" attacks.

The purpose of the work is related to the development of a methodology that allows for automated management of multi-level access to information and program resources of educational organizations to increase the level of protection.

The main tasks to be solved in the work: analysis of the characteristics of information processing in the system, development of a technique that allows for the design of an automated system aimed at distinguishing access to information and software, the development of a software module that allows you to manage the database of an automated system of an educational company to ensure

Areas of access to strategically important information are given. The algorithm of work of the administrator of educational institution with system is presented. The scheme of tables of the database of the institution is presented. Viewing the table of incidents of violation of the rules of work with information and software of the system is available only to employees of the Department of information security of educational institution and only in the reading mode.

Changing the data of the table is not available to the staff of the institution or the system itself. The system produces only "additional records" of new table tuples. When logging in, the employee must enter the appropriate identifier and authenticator (in the example provided, the authenticator is the password).

The interrelations between the structural units of the educational institution for the subsequent analysis of information flows are revealed. On the basis of the identified relationships, a structural model of automated information flow management based on the principles of information and software protection of the system is developed..

Keywords:

Information security, automated workplace.

ACM Computing Classification System:

Security services, Human computer interaction , Information storage systems

Introduction

Educational institutions are increasingly faced with the fact that there are problems of violations of the protection parameters of information and software, they are determined by the influence of unauthorized actions of legal users – employees of the institution.

At the moment, there is a set of systems to prevent unauthorized use of information and software, but none of them provides comprehensive protection of resources.

The need for comprehensive protection of information and software resources determines the development of an integrated system that is integrated into the main information environment of educational institutions, its main focus is related to the prevention of the maximum number of types of "insider" attacks.

The complexity in the structure of information and software leads to the fact that the developed automated system for the differentiation of access to information and software has a complex structure [1].

The purpose of the work is related to the development of a methodology that allows for automated management of multi-level access to information and program resources of educational organizations to increase the level of protection.

In the paper we consider these problems:

1. Implementation of analysis of information processing characteristics in the system.
2. Development of a methodology that allows for the design of an automated system aimed at the differentiation of access to information and software.
3. Development of a software module that allows you to manage the database of the automated system of the educational company to ensure the control of the reliability of its paper.

1 Analysis of the main functions of the automated system of differentiation of access to information and software of the educational institution according to the method of multi-level access management

The main focus of the developed system is related to the provision of automated control of multi-level access to information and software resources.

Information resources should be classified according to three criteria:

- Access level;
- Relation to a particular unit of the institution (developing units, secondary services, etc.));
- Presentation form (text, graphic documents, etc.).

Multi-level access management should be carried out taking into account all the criteria of classification of information resources, taking into account the study of information flows of educational institutions and the features of joint access of employees to information and program resources.

Functions of management of multilevel access to information resources are information and management.

The information function is a set of two mechanisms [2]:

1) the mechanism of control of the state of subjects of protection (information and program resources;
2) the mechanism of informing on incidents of violation of rules of work with information and software of educational institution.

The control function is the application of the basic algorithm of differentiation of access to information and software resources. Also the control functions include:

- accounting of information and software resources;
- Resource storage;
- The results of the requested staff resources;
- The results of media;
- Password issuance;
- Keys.

At the same time, the generation of passwords, keys, support of access control, as well as control over the process of processing of information resources is entrusted to the staff of information security units of the educational institution.

Also, management is necessary at all stages of implementation of the system, including at the stage of creation, at the stage of operation.

2 Development of a structured model for the company with the condition of protection of its information and software

To create a structural model of the educational institution management it is necessary to identify the relationships of the structural units of the institution in terms of information and software protection management [3, 4].

Figure 1 shows the areas of access to strategic information.

The upper level includes personal data of students, office management system and accounting programs, all this is designed to maintain records of students, their parents, teachers (employees) for the operational management of educational institutions.

The highest levels of access ("red") are those structural units of the educational institution, which are directly involved in the processing of information about students and staff of the institution, as well as the development of projects for the educational process.

The work of these units is connected, so it is quite difficult to define the border of their access to information about the project.

In order to achieve the main development goals, it is necessary that the area of access of research and development units involved in the same project is as much as possible coincided.

Figure 1 shows that the top level also includes the departments of standardization, quality control and technical documentation control.

These departments also have access to strategic information about the project being developed, so they need to be given special attention.

The staff of the General Department do not have access to any strategically important information resources of the educational institution system, so they belong to the "green" level of access.

Access Manager-is an additional module, which is used to organize the access of the participant to the protected information, the main purpose of this module is authorization and authentication of the employee.

To start the operation of this module, it is necessary to build a database of members of the organization, which will be entered all the names, information about access marks, biometric characteristics of the employee, password.

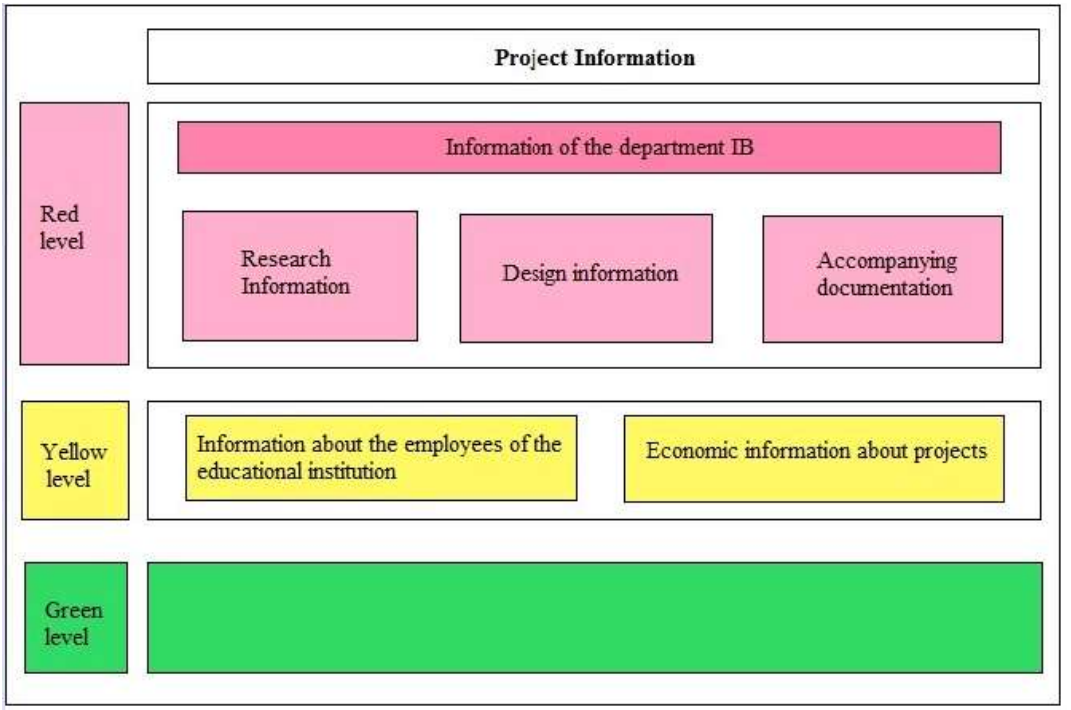


Fig.1. Project information with access areas

3 Algorithm of administrator work of educational institution with the automated system of differentiation of access to information and software

Figure 2 shows the algorithm of the administrator of the educational institution with the system. A member of the organization provides the following information as an identification, to choose from:

- Password;
- Material carrier;
- Employee biometric data.

Then the processes take place:

- Identification;
- Authentication;
- Authorization of employees.

Control is transferred to the beginning of the algorithm, if the process was unsuccessful, otherwise the participant of the organization is given a label L, which displays at what level of access is a particular employee. There are three levels of green, yellow and red access in the system.

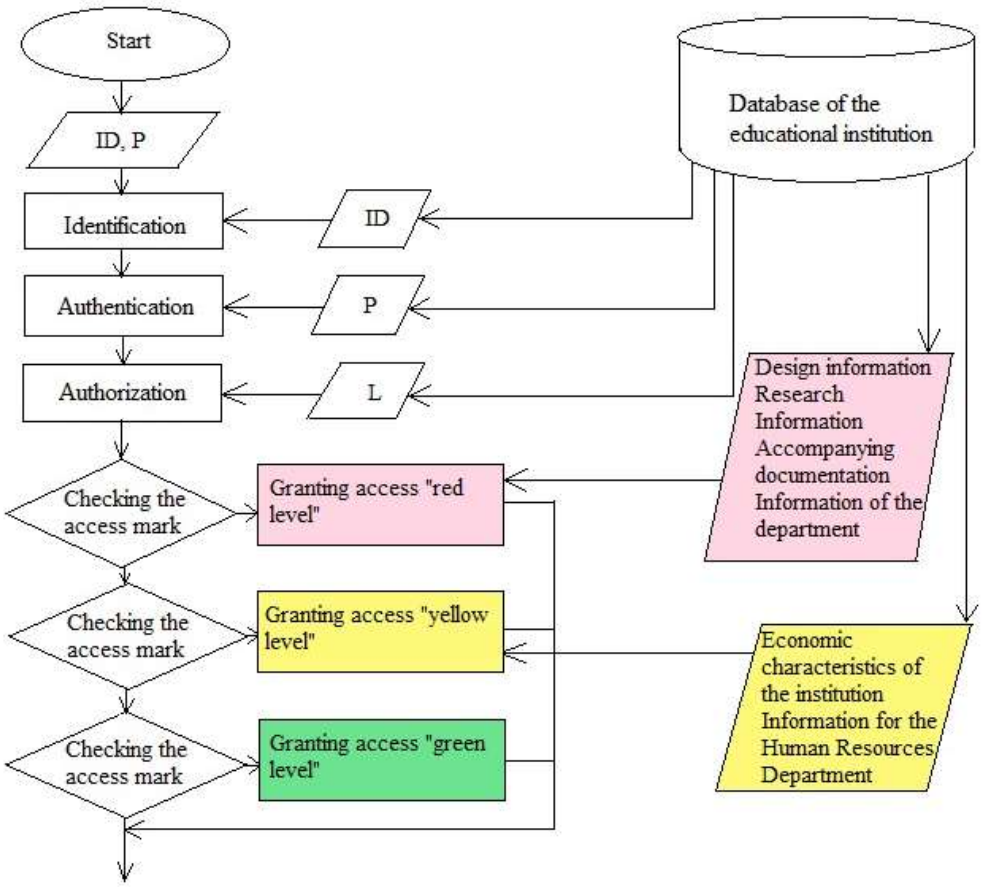


Fig.2. Algorithmic scheme of the organization participant's work with the system

4 The database structure of the organisation

To implement the above methodology, there is a need to create a database, which will provide information about both the subjects and the objects of protection. Figure 3 shows the schema of the institution's database tables.

5 The main components of the system

The main components of the system are shown in figure 4. The main task solved by this subsystem of differentiation of access of employees to information and software resources of the system is to ensure compliance with the differentiation of access through the use of organizational, legal, technical (hardware and software), physical protection measures [23].

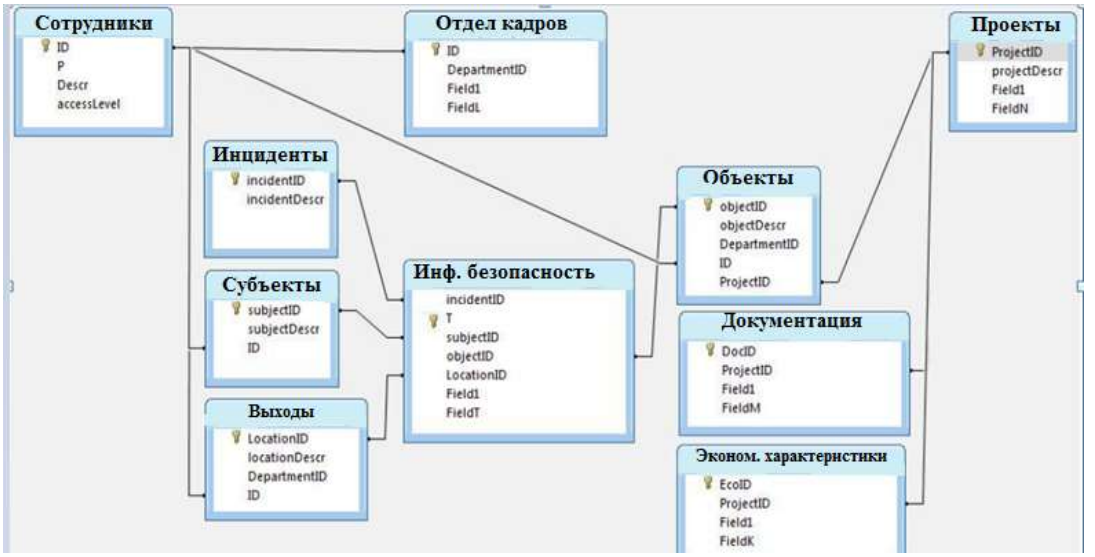


Fig.3. Diagram of database tables of an educational institution

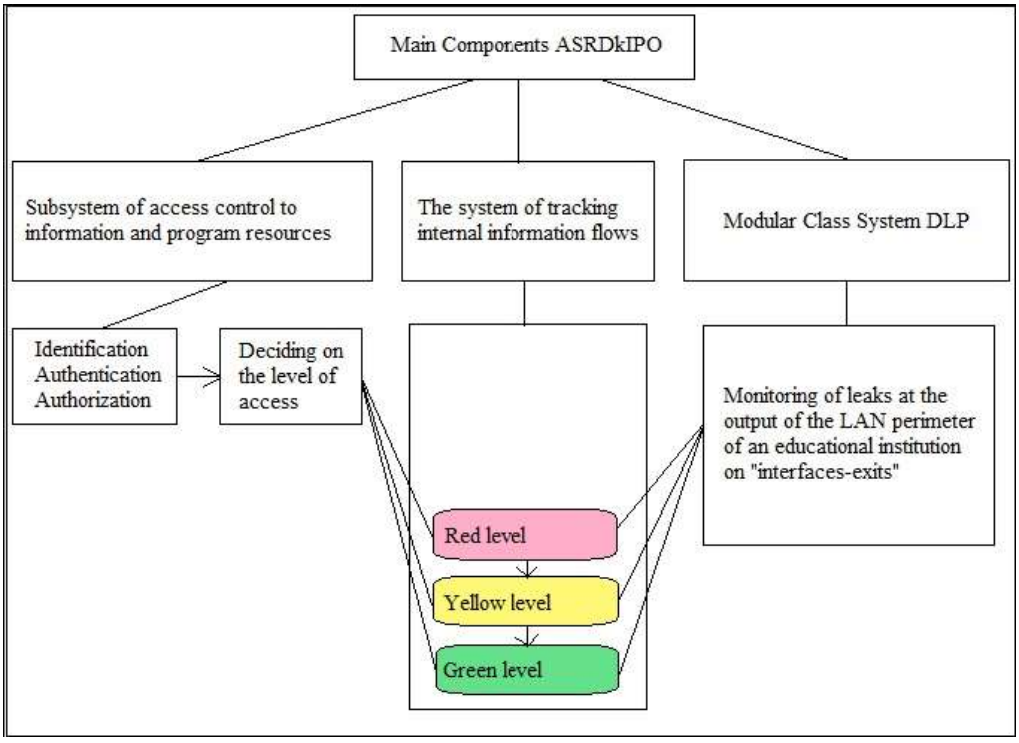


Fig.4. Main components of the system

6 Forms of automatic workplace of administrator of the educational institution designated

Form of work with audit log-table of incidents of violation of rules of work with information and software of the system. An example of the form of work with the audit log is shown in figure 5.

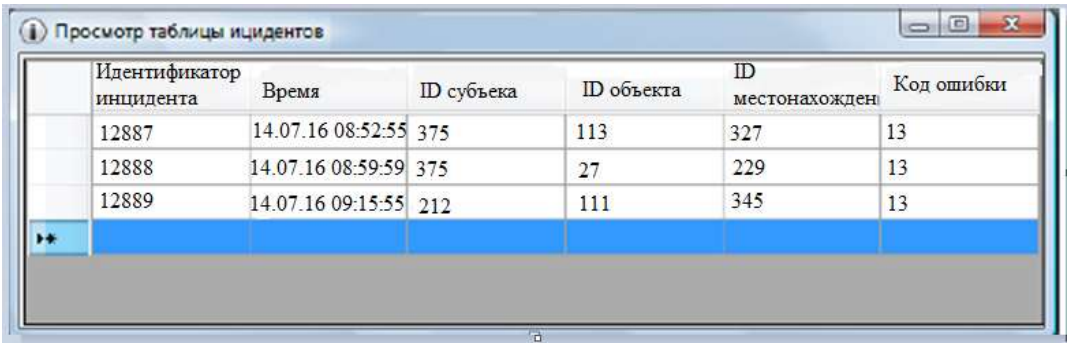


Fig.5. Example of the audit log form

All fields in the table are read-only. The system automatically writes to the table.

Viewing the table of incidents of violation of the rules of work with information and software of the system is available only to employees of the Department of information security of educational institution and only in the reading mode [5, 6].

Changing the data of the table is not available to the staff of the institution or the system itself. The system produces only "additional records" of new table tuples.

When logging in, the employee must enter the appropriate identifier and authenticator (in the example provided, the authenticator is the password). The method of password authentication can also be made customizable (to regulate the input set of characters, the number of characters, the waiting period after a failed authentication, etc.) menu Form of work with the database of the employee of the information security Department of the enterprise. An example of the form is shown in figure 6.

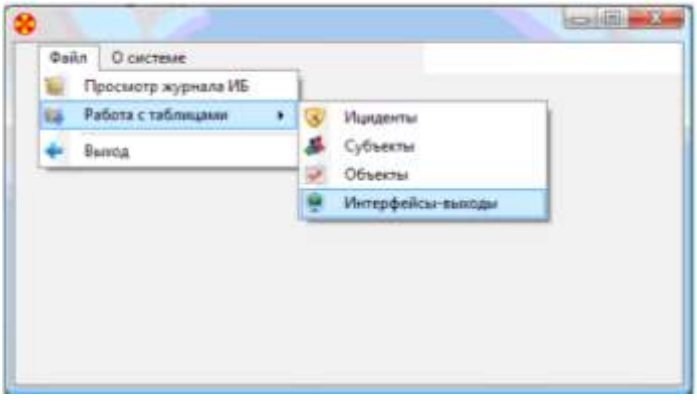


Fig.6. Example of the form of work of the information security Department employees

As shown in the figure, an employee of the information security Department of the educational institution can cause:

- The form of work with the audit log (File - view log IB);
- The form of a table of descriptions of incidents of information relations (File – table Incidents);
- Form of work with the table of description of subjects of information relations (File - work with tables-Subjects);
- The form of work with the table of objects of information relations (File - work with tables-Objects);
- The form of work with the table "Interfaces-inputs" (File – work with tables – Interfaces-outputs).

As shown in figure 7, the Agency's human resources staff can call:

Form of work with the table "personnel Department of educational institutions" (file Employees).

Types of forms of work of employees of the institution were designed as one basic form. For the necessary differentiation of the property "Visible" of the relevant fields in the menu equivalent to "false" figure 7 shows the properties window for the menu fields of the form.

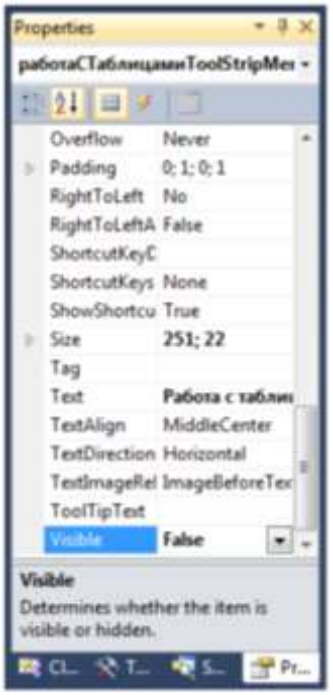


Fig.7. Properties of the fields of the menu of the form of work of employees of educational institution

The type of the basic form of work of employees of educational institution with information and program resources of system is presented in figure 8.

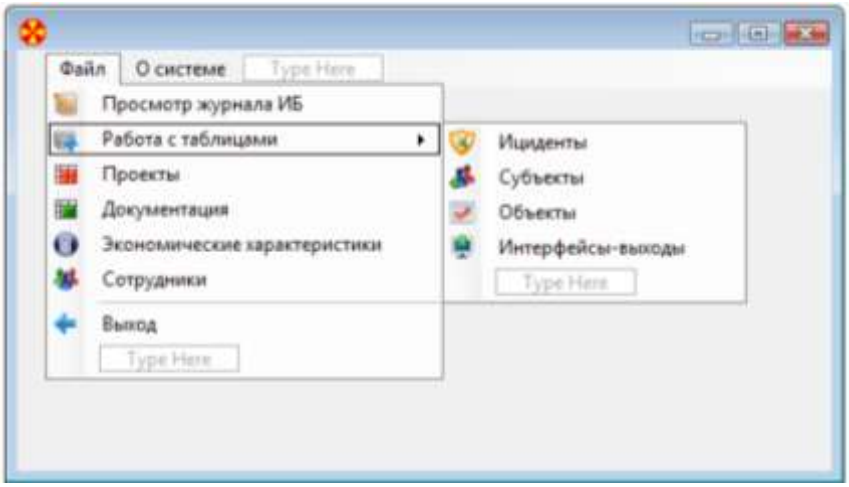


Fig.8. Form view when developing a project in the Microsoft Visual Studio 2010 programming environment

Conclusion

In order to ensure the proper level of protection of information and software systems of educational institutions developed a technique of automated management of access to strategic information and software resources for the administrator of the institution [7].

The interrelations between the structural units of the educational institution for the subsequent analysis of information flows are revealed. On the basis of the identified relationships, a structural model of automated information flow management based on the principles of information and software protection of the system is developed.

The basic modules of the automated system of differentiation of access to information and software are designed.

On the basis of the analysis in the design process developed methods to improve existing solutions, taking into account the identified main advantages and disadvantages of similar modules of domestic and foreign manufacturers.

The algorithm of management of differentiation of access of employees of educational institution to information structures including processes of identification, authentication and authorization of employees [7] is created.

According to the created algorithm, the database of the educational institution is designed, the variant of differentiation of access to the database tables for employees of structural units is presented.

The program complex of work of the administrator of the automated system with the database of employees of the institution is developed.

References

[1] Preobrazhensky Yu. P. Information security - challenges of the modern world / / Bulletin of the Voronezh Institute of high technologies. 2017. № 2 (21). P. 60-63.

[2] Marichev V. A., Lyubimov V. I., Preobrazhenskiy Yu. P .Issues of social engineering in corporate information security / / Bulletin of the Voronezh Institute of high technologies. 2017. № 2 (21). P. 64-67.

[3] Lvovich I. Ya., Voronov A. A., Preobrazhenskiy Yu. P. factors of threat to economic security of the state. Information and security. 2006. Vol. 9. No. 1. Pp. 36-39.

[4] Voronov A. A., Lvovich I. Ya., Preobrazhenskiy Yu. P., Voronov V. A. provision of risk management system in case of information security threats // Information and security. 2006. Vol. 9. No. 2. Pp. 8-11.

[5] Preobrazhenskiy Y. P., Preobrazhenskaya N.S. Lvovich I. Ya. Some aspects of Informatization of educational institutions and the development of media competence of teachers and managers // Herald of the Voronezh state technical University. 2013. Vol. 9. No. 5-2. P. 134-136.

[6] Chernikov S. Yu., Korolkov R. V. use of system analysis in the management of organizations / / Modeling, optimization and information technologies. 2014. No. 2 (5). P. 16.

[7] Gusev M. E., Zhigalkina T. A., Khorseva O. V., Kruglyakova E. A., Preobrazhenskiy A. P. Problems of training of specialists in the field of Informatization of education // Bulletin of Moscow city pedagogical University. Series: Informatics and Informatization of education. 2006. No. 7. P. 223.

Prof. Yakov Lvovich
Doctor of Sciences (Engineering), professor,
Voronezh Institute of High Technologies, Russia
office@vivt.ru

Prof. Vera Kostrova
Doctor of Sciences (Engineering), professor,
Voronezh state technical university, Russia
V_kostrova@mail.ru

Ing. Juraj Štefanovič, PhD.
Faculty of Informatics, Pan-European University, Bratislava, Slovakia
juraj.stefanovic@paneurouni.com

DEVELOPMENT OF AN AUTOMATED WORKPLACE FOR THE PC COMPONENTS WAREHOUSE

Igor Lvovich, Yuriy Preobrazhenskiy, Eugen Ružický

Abstract:

Today, automated workplaces can be observed in a variety of fields. In this work, the development of an automated workplace of the call-center operator for technological diagnostics, repair and maintenance of the PC system unit was made. The purpose of the work is to develop an automated workplace for the operator to manage the warehouse of components for the PC. When developing a model of an automated workplace, it is necessary that the program performs a number of functions that will help to increase the efficiency of the head of the organization: input of initial information and filling in the database; correction and storage of data; presentation of data in a convenient form; issuance of reference data. Taking this into account, the structure of the automated workplace software is formed. One of the parts of the successful work of the call - center and guarantee of long-term profitable relationships with customers is high-quality and efficient service. Quality and efficiency of service is formed: from the availability of a single database; of highly qualified personel; from the established system of processing applications; from the availability of high quality material and technical base; of the profit of the vehicle and mobile communications. The service center is the principal operations: receiving requests from customers; storage of applications in the information base; create progress reports for clients; monitoring progress and contributing to the database on the current work. For the distribution of components in the warehouse, we were based on the solution of the linear programming problem. To determine the distribution of PC components in warehouses, the algorithm for calculating the roots of the objective function under these restrictions was considered.

Keywords:

Warehouse, automated workplace, information system.

ACM Computing Classification System:

Computing and business, Service-oriented architectures.
Interaction design process and methods

Introduction

The basics of information and other software related to their use as a tool occupy an important place in the course of the company. Their application makes it possible to reduce the time required for processing customers ' orders, and, consequently, the speed of service with customers in the enterprise as a whole.

Of course, in order to identify all possible abilities, without exception, which carries the use of databases, it is necessary to use a complex of software and hardware as detailed as possible to the set tasks. Due to this, at the present time there is a great need for computer programs that would support and coordinate the process [1, 2].

Today, automated workplaces can be observed in a variety of fields. In this work, the development of an automated workplace of the call-center operator for technological diagnostics, repair and maintenance of the PC system unit was made.

The purpose of the work is to develop an automated workplace for the operator to manage the warehouse of components for the PC.
Work tasks:

1) workplace automation analysis;
2) analysis of the operator's activity in warehouse management;
3) development of a model of an automated workplace for warehouse management;
4) development of the operator's arm functioning algorithm for warehouse management;
5) development of the operator's arm user interface for warehouse management;
6) development of the operator's arm program for warehouse management;
7) realization of the operator's arm model for warehouse management, repair and maintenance of the PC system unit.

1 Conceptual modeling of the automated workplace of the operator on warehouse management of PC components

The operator's workstation of the call-center needs to be provided with the following tasks:

- consultation on required data;
- automatic processing and formation of waybills;
- routing for scheduled jobs and the control of the route;
- payroll and the cost of transportation.

Making analysis of the listed responsibilities of the operator and the tasks which are solved with the help of automation. In this paper, a program that will allow:

- shaping directions in automatic mode;
- keep a log of the work of drivers and vehicles;
- issuance and storage of reference data;
- creating and updating client database.

To solve the tasks, you need the source, output and intermediate data. To create a client database, source data:

- type of work;
- the name of the head;
- phone number;
- address for correspondence;
- legal form;
- payment details;
- long-term plan.

These data are needed to implement the automated workplace function, can be represented in the form of a relational data model [3,4].

Data redundancy is the repetition of values in tables or attributes.

Due to the compilation of tables for different sources, there is a data inconsistency.

Data computability is the significance of individual attributes can be determined according to known values.

Information is almost impossible to lay out on minor and elementary relations, and there is no connection between the information, it is permissible to move to the compilation of the file texture.

The structure of the file for saving "Clients" information is shown in table 1.

Table 1 - Structure of the file for storing "Clients" information

Field name	**Field type**	**Field length**
Legal form	Line	100
Address for correspondence	Line	100
Shipping and payment details	Line	100
Phone number	Line	11
The name of the head	Line	30
Type of work	Number	30
Perspective (date and order quantity)	Line	30

On the basis of this file structure, the necessary documents are created. The program sets the necessary filtering and sorting settings depending on the type of document. Also, certain arithmetic operations are carried out with the data. The output is projected as raw information that has been processed or in the original form.

When developing a model of an automated workplace, it is necessary that the program performs a number of functions that will help to increase the efficiency of the head of the organization:

- input of initial information and filling in the database;
- correction and storage of data;
- representation of the data in a convenient form;
- the results of reference data.

Taking this into account, the structure of the automated workplace is formed (figure 1).

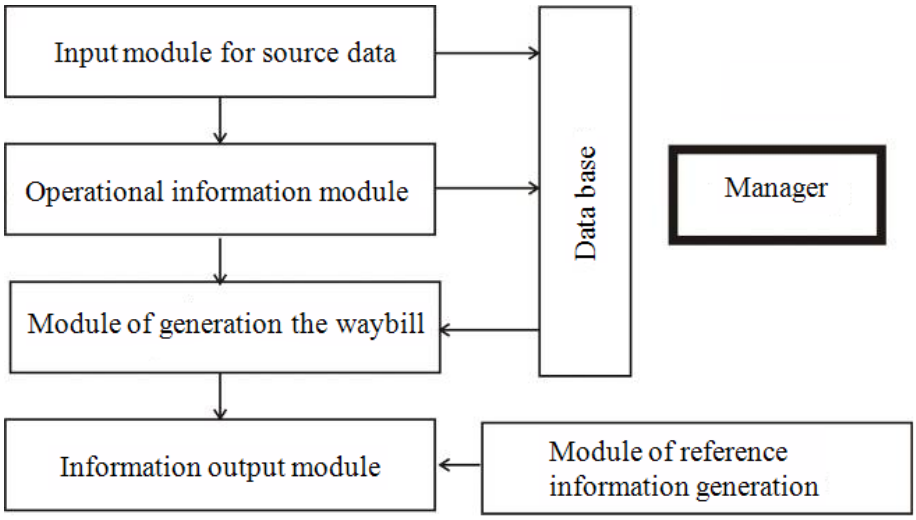


Fig.1. Structure of the automated workplace

The structure of this program consists of seven modules:

* the call centre operator or the operator;
* initial information input module;
* module for the accounting of timely information;
* module for editing waybills;
* information base;
* module for generating reference information;
* output module.

For specific tasks, there are a large number of different algorithms for their resolution. Finding the necessary method is subject to additional conditions that apply to these methods. The procedure for forming a method is associated with the recording method in a particular language.

The algorithm of the automated workplace model can be different. It uses all types of algorithms: linear, cyclic, branching. All this is necessary for accurate data processing and creating the necessary reports.

The linearity of the algorithm is the sequence of actions in the basic module of the program. Branching algorithm is needed to extract information from files. That is, when you press the desired key, a process is performed, which is described in the event according to the pressing of each key.

The method of this program uses opposite connections, which allow the operator to choose the necessary approach to continue the program.

The call-center operator executes the project management. In the process, it contains the necessary modules and ensures their correct interaction.

The advantage of this method is the ease and probability of structure and connection of specific elements to be clearly reflected.

When developing this program, it is important to pay special attention to the problem of accessibility, clarity and convenience of dialogue between the PC and the operator.

The program that creates a model of a call center specialist workstation is necessary to perform on the principle of modularity. This makes it more versatile and allows for upgrades as needed [5, 6].

For the convenience of the operator you need to display the necessary tips and buttons.

One of the parts of the successful work of the call - center and guarantee of long-term profitable relationships with customers is high-quality and efficient service.

Quality and efficiency of service is formed:

- from the availability of a single database;
- of highly qualified personnel;
- from the established system of processing applications;
- from the availability of high quality material and technical base;
- of the profitability of the vehicle and mobile communications.

The service center is the principal operations:

- receiving requests from customers;
- storage of applications in the information base [7];
- create progress reports for clients;
- monitoring progress and contributing to the database on the current work.

2 Description of the linear programming problem

To determine the distribution of PC components in warehouses, consider the algorithm for calculating the roots of the objective function under these restrictions.

Variables that are included in only one equation of the constraint system with a coefficient of 1 are called basic variables (for our problem these are variables). The other variables are called free variables. Then equating the basis variables to the corresponding right-hand parts of the constraints, and the free variables to zero , we obtain a reference plan defined by a system of unit vectors, which form the basis of the m-dimensional space.

For the convenience of calculations in the simplex method, all data on the problem are recorded in the simplex table:

Table 2 - Simplex table

x_δ	c_δ	B	c_1	c_2	…	c_k	…	c_j	…	c_n
			x_1	x_2	…	x_k	…	x_j	…	x_n
$x_{\delta 1}$	$c_{\delta 1}$	b_1	a_{11}	a_{12}	…	a_{1k}	…	a_{1j}	…	a_{1n}
…	…	…	…	…	…	…	…	…	…	…
$x_{\delta s}$	$c_{\delta s}$	b_s	a_{s1}	a_{s2}	…	a_{sk}	…	a_{sj}	…	a_{sn}
…	…	…	…	…	…	…	…	…	…	…
$x_{\delta i}$	$c_{\delta i}$	b_i	a_{i1}	a_{i2}	…	a_{ik}	…	a_{ij}	…	a_{in}
…	…	…	…	…	…	…	…	…	…	…
$x_{\delta m}$	$c_{\delta m}$	b_m	a_{m1}	a_{m2}	…	a_{mk}	…	a_{mj}	…	a_{mn}
		F	Δ_1	Δ_2	…	Δ_k	…	Δ_j	…	Δ_n

In the top row of the table coefficients are made for all variables in the objective function.

In the first column of the table, the basic variables are introduced in the order in which they come into the concept of restrictions, in the second – the coefficients of the objective function for the basic variables, in the third – the right shares of all restrictions, in the subsequent columns – the coefficients for certain unstable constraints in the concept.

In the lower line of the table, estimates are made according to any variable, which are then characterized by:

It is obvious that for the basis unstable estimates are zero.

At any iteration of the simplex method, the conclusion from the basis of some other unstable variable and the inclusion of another variable with the proper recalculation of the table components are carried out. Before solving the problem it should lead to the canonical figure.

3 Using the simplex method to implement a model of the operator's automated workplace for the management of the components PC warehouse

The organization consists of 4 branches, each of which has its own warehouse. For example, there are 3 types of goods that are needed in these warehouses. The number of goods of each type, which should be on average in the warehouses is given in the table:

Table 3 - Example of using the simplex method

Type of goods	1 warehouse	2 warehouse	2 warehouse	3 warehouse	The total number on the basis of wholesale
I	3	2	2	0	20
II	1	1	3	2	37
III	1	1	0	4	30
Profit from the sale of 1 PC, rub	9	6	11	6	

The table shows the total number of goods of each type on the wholesale basis and profit from the sale of one unit of goods.

Determine how many goods of each type should be in warehouses to maximize the profit from the sale of goods.

The final version of the system of equations:

x 0 = 137-7 / 8 x 2-17 / 8 x 5-9 / 4x6-3/8 x 7
x 3 = 7-1/16 x 2+1/16 x 5-3/8x6+3/16x7
x 4 = 7-3/32 x 2+3/32 x 5-1/16 x 6-7/32 x 7
x 1 = 2-5/8 x 2-3/8 x 5+1/4x6-1/8 x 7

The optimal plan can be recorded as follows:

x3 = 7
x4 = 7
x1 = 2
F (X) = 137.

4 Software implementation of the Simplex method

As a result of the analysis of the task and trading activity of the enterprise, a software implementation was developed for the solution using the simplex method.

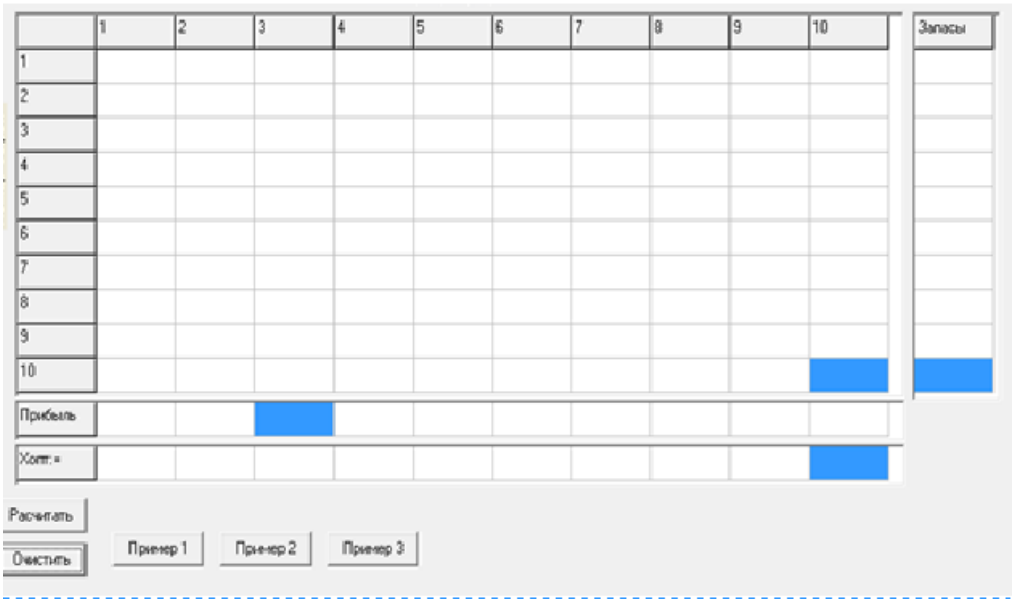


Fig.2. Window interface of the developed program

For the problem discussed in detail in the example above, the solution using the program will look like this:

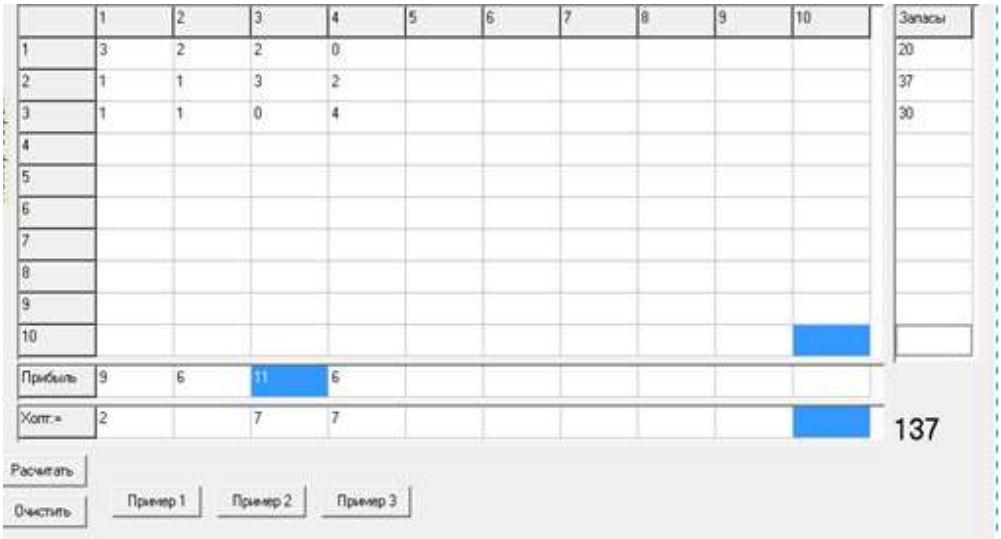


Fig.3. Example of solving the problem using the simplex method

After receiving the result, we can draw the appropriate conclusions. With a given number of PC components of each type, the maximum revenue from the sale of these products will be 137.

Let us consider a few more examples and solve them with the help of the developed program. With the given parameters, to determine the maximum income from the sale of PC components:

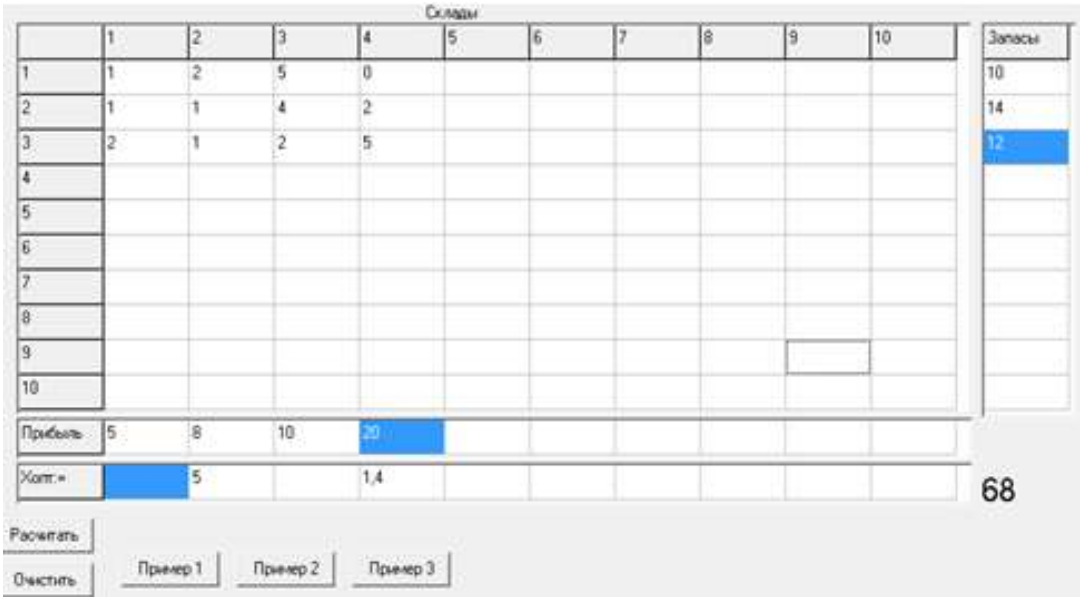


Fig.4 An example of solving the problem by simplex method.

The maximum income from a given amount of goods in this problem will be 2860 rubles.

Testing of the developed program showed that the software completely copes with the tasks assigned to it, finding the maximum in functions under the given restrictions.

Table 4 - Example of ordering of goods by the organization

Type of goods	**1 warehouse**	**2 warehouse**	**2 warehouse**	**3 warehouse**	**The total number on the basis of wholesale**
TV BBK	0	0	0	10	170
Monitor LG	0	30	10	0	200
Camera-re-corder	6	0	0	0	120
Profit from the sale of 1 PC rub.	5400	2500	2500	3700	

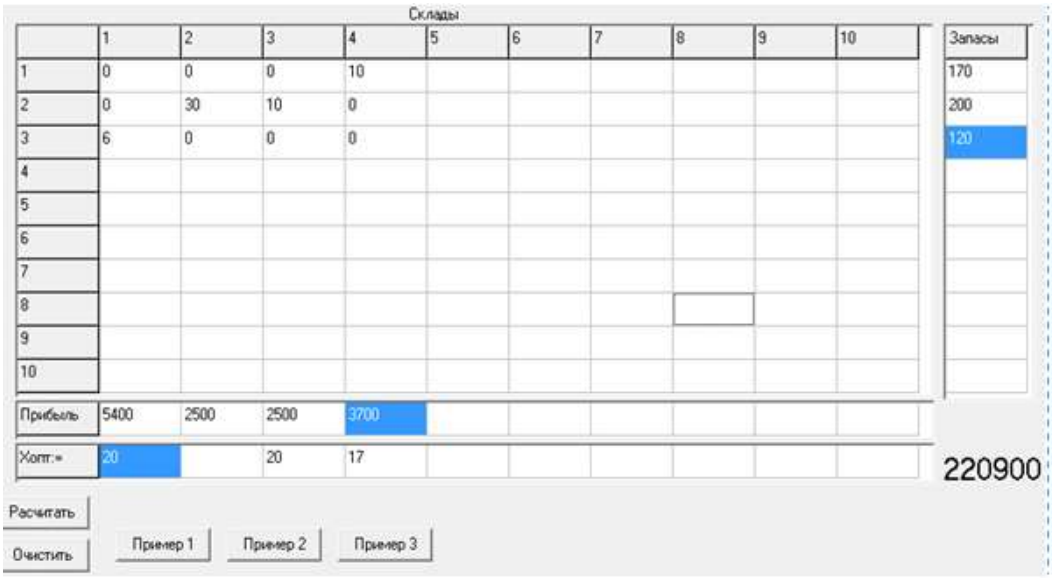


Fig.5. Determination of income from the ordered goods

The result shows that the maximum profit from the order will be 220900 rubles.
Let's analyze one more order:

Table 5 – Example of ordering of goods by the organization

Type of goods	**1 warehouse**	**2 warehouse**	**2 warehouse**	**3 warehouse**	**The total number on the basis of wholesale**
Keyboard BTC	50	0	0	20	300
Монитор LG	0	0	20	0	200
Camera-re-corder	0	40	0	0	120
Profit from the sale of 1 PC rub.	300	3200	2500	300	

Fill in the appropriate table in the developed program:

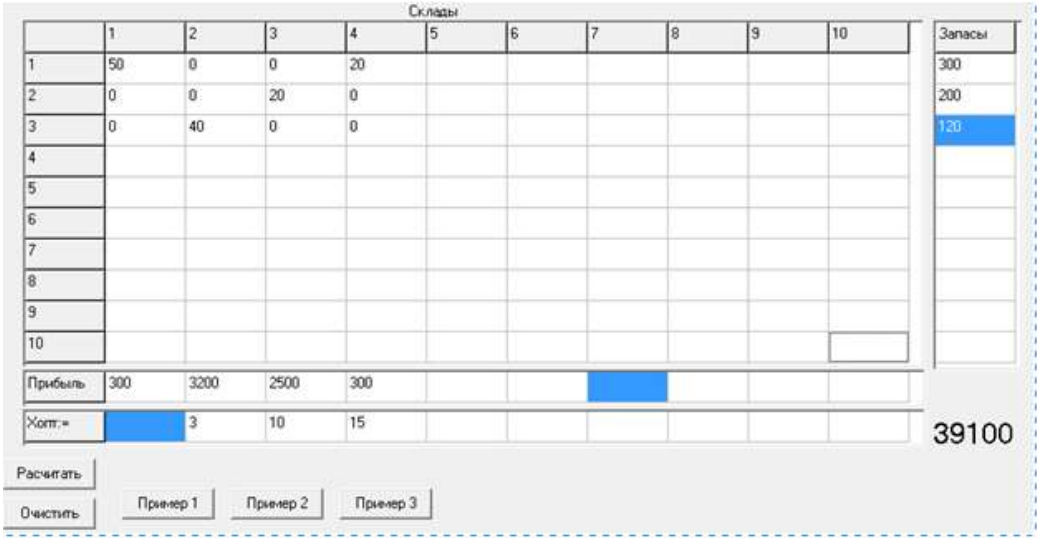


Fig.6. Determination of income from a given product

The income from this order was 39100 rubles.

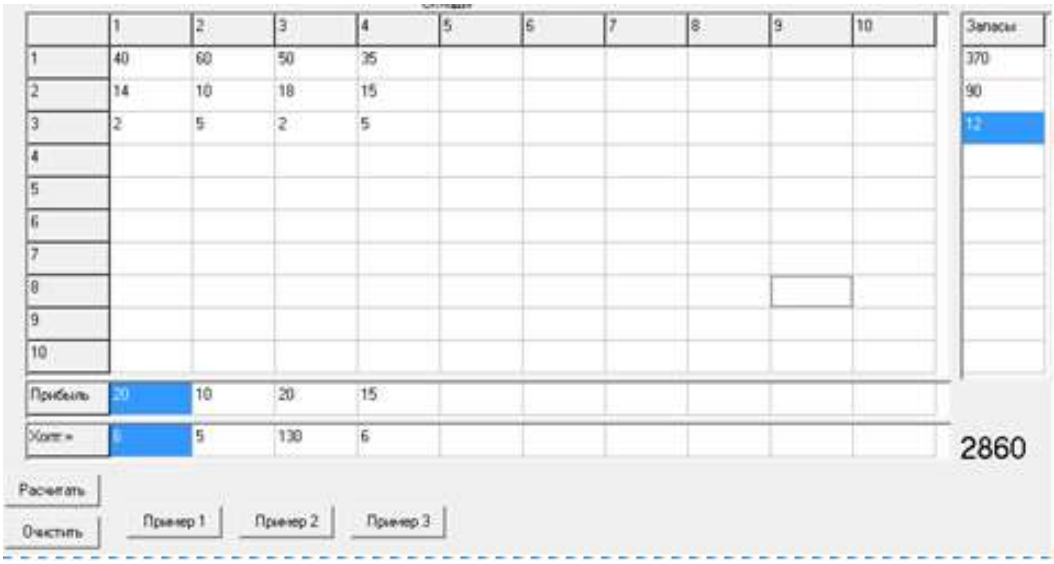


Fig.7. An example of solving the simplex problem

Now, from abstract numbers, go directly to the problem that will be solved in the trading organization with the help of the developed program.

Table 6 – Example of ordering of goods by the organization

Type of goods	**1 warehouse**	**2 warehouse**	**2 warehouse**	**3 warehouse**	**The total number on the basis of wholesale**
Keyboard BTC	50	0	0	20	300
Монитор LG	0	0	20	0	200
Camera-re-corder	0	40	0	0	120
Profit from the sale of 1 PC rub.	300	3200	2500	300	

We will enter the data into the developed program.

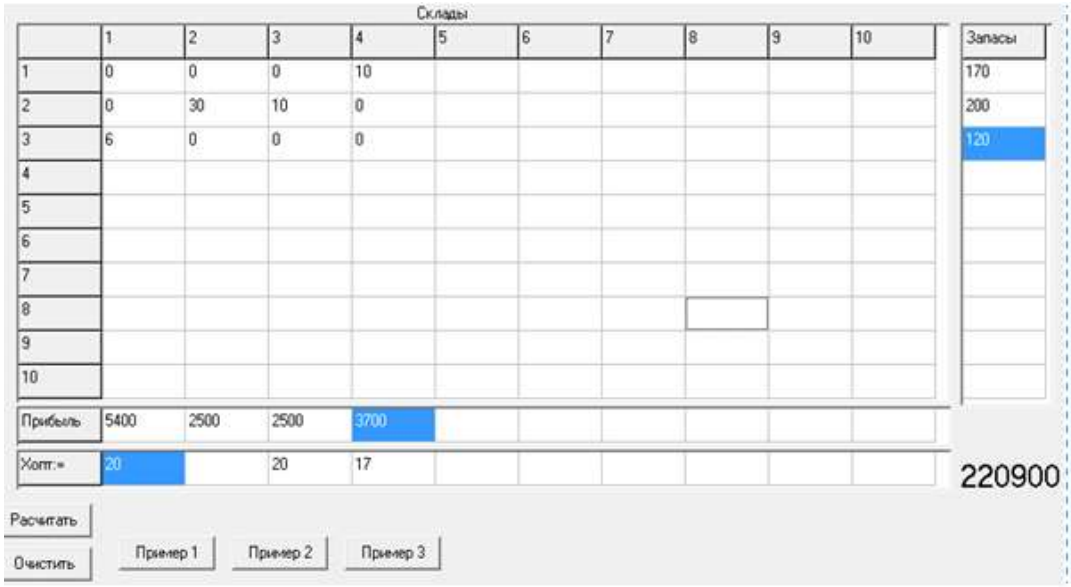


Fig.8. Determination of income from the ordered goods

The result shows that the maximum profit from the order will be 220900 rubles.

Let's analyze one more order:

Table 7 – Example of ordering of goods by the organization

Type of goods	**1 warehouse**	**2 warehouse**	**2 warehouse**	**3 warehouse**	**The total number on the basis of wholesale**
Keyboard BTC	50	0	0	20	300
Монитор LG	0	0	20	0	200
Camera-re-corder	0	40	0	0	120
Profit from the sale of 1 PC rub.	300	3200	2500	300	

Fill in the appropriate table in the developed program:

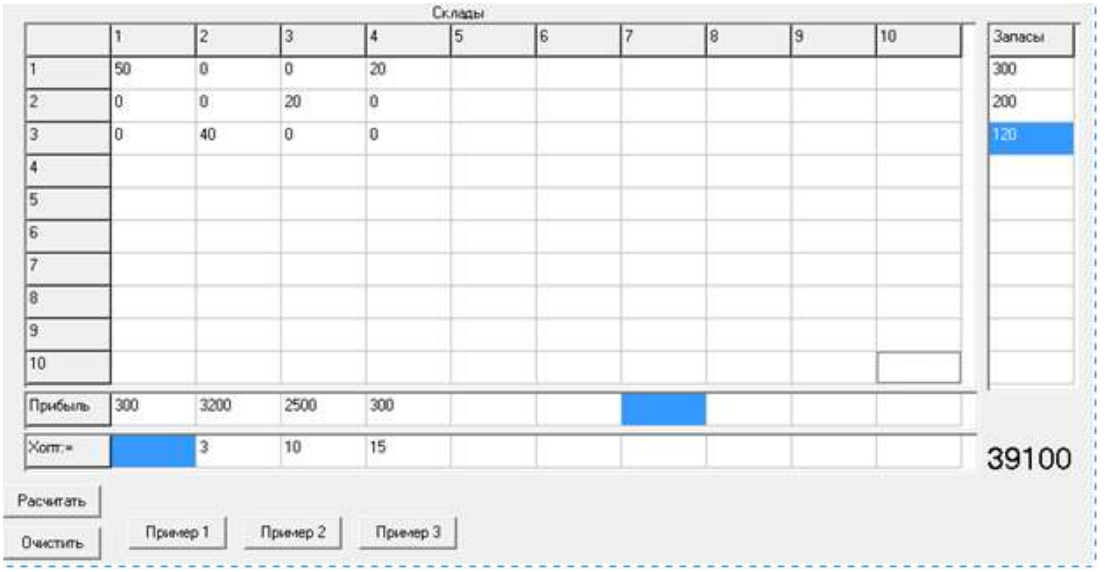


Fig.9. Determination of income from a given product

The income from this order was 39100 rubles.

Conclusion

The placement of goods in the warehouse plays a key role in the optimization of warehouse processes. The quality and speed of selection depends on how the goods are placed in the warehouse.

In this paper, the problem of goods placement in warehouses was presented in the form of a linear programming problem and solved using the simplex method. For the presented problem the program was made. The use of the created program at the enterprise allowed to expand the method of storage used in the organization in a warehouse.

References

[1] Moskalchuk Yu. I., Naumova E. G., Kiseleva E. V. problems of optimization of innovative processes in organizations // Modeling, optimization and information technologies. 2013. No. 2. P. 10.

[2] Preobrazhenskiy Y. P., Panevin, Y. R. The formulation and classification of problems of optimal control of production objects // Bulletin of Voronezh state technical University. 2010. Vol.6. No. 5. P. 99-102.

[3] Zemlyanukhina N.S. With. On the application of information technology in management // Advances in modern natural science. 2012. No. 6. P. 106-107.

[4] Preobrazhenskiy Yu. Evaluation of the effectiveness of the system of intellectual support of decision-making // Bulletin of the Voronezh Institute of high technologies. 2009. No. 5. P. 116-119.

[5] Penkov P. V. Expert methods of control systems improvement // Bulletin of the Voronezh Institute of high technologies. 2012. No. 9. P. 108-110.

[6] Samoylova U. A. on some characteristics of enterprise management // Bulletin of the Voronezh Institute of high technologies. 2014. No. 12. P. 176-179.

[7] Korolkov R. V. Controlling in the trade organization / / Bulletin of the Voronezh Institute of high technologies. 2013. No. 10. P. 287-290.

Prof. Igor Lvovich
Doctor of Sciences (Engineering), professor,
Voronezh Institute of High Technologies, Russia
office@vivt.ru

Assoc. Prof. Yuriy Preobrazhenskiy
Candidate of Sciences (Engineering), associate professor,
Voronezh Institute of High Technologies, Russia
office@vivt.ru

Assoc. Prof. Eugen Ružický, PhD.
Faculty of Informatics, Pan-European University, Bratislava, Slovakia
eugen.ruzicky@paneurouni.com

INTELLECTUAL SUPPORT OF MEDICAL DECISIONS IN SELECTION OF MEDICAMENTAL TREATMENT BASED ON THE DUAL-MODEL APPROACH TO FORECASTING AND OPTIMIZATION OF ITS EFFECTIVENESS

ИНТЕЛЛЕКТУАЛЬНАЯ ПОДДЕРЖКА ВРАЧЕБНЫХ РЕШЕНИЙ ПРИ ПОДБОРЕ МЕДИКАМЕНТОЗНОГО ЛЕЧЕНИЯ НА ОСНОВЕ ДУАЛЬНОМОДЕЛЬНОГО ПОДХОДА К ПРОГНОЗИРОВАНИЮ И ОПТИМИЗАЦИИ ЕГО РЕЗУЛЬТАТИВНОСТИ

Elena Gafanovich, Artem Lvovich
Елена Гафанович, Артем Львович

Abstract:

The article presents the dual-model approach aimed at intellectual support of medical decisions in selection of medicamental treatment. The duality is determined by the two ways of use of the neural network model, linking the dynamics of the treatment effectiveness with the parameters characterizing a patient's initial condition and a medical prescription. It is demonstrated that the first method is based on the replacement of real performance assessments in control points with prognosis evaluation in medicamental treatment correction. The second method allows to include expert evaluation of a doctor in the decision-making process with an orientation to the optimal doses of medical drugs, calculated on the basis of a mathematical model. The combination of these methods ensures an increase in the effectiveness of the decisions made by reducing the number of control points and selecting a more appropriate medicamental treatment for a patient.

Keywords:

Medicamental treatment, dual-model approach, forecasting, optimization, neural network modeling, expert evaluation.

ACM Computing Classification System:

Artificial Intelligence – Applications and Expert Systems
Artificial Intelligence – Knowledge Representation Formalisms and Methods
Simulation and Modeling – Applications

Абстракт:

В статье рассматривается дуальномодельный подход, направленный на интеллектуальную поддержку врачебных решений при подборе медикаментозного лечения. Дуальность определяется двумя способами использования нейросетевой модели, связывающей динамику результативности лечения с параметрами, которые характеризуют исходное состояние больного, и назначениями врача.

Показано, что первый способ основан на замене реальных оценок результативности в контрольных точках прогнозными оценками при коррекции медикаментозного лечения. Второй способ позволяет включить экспертные оценки врача в процесс принятия решений с ориентацией на оптимальные дозы лекарственных препаратов, вычисленные на основе математической модели. Совмещение этих способов обеспечивает повышение эффективности принимаемых решений за счет сокращения числа контрольных точек и подборе более адекватного состоянию больного варианта медикаментозного лечения.

Ключевые слова:

Медикаментозное лечение, дуально-модельный подход, прогнозирования, оптимизация, нейросетевое моделирование, экспертное оценивание..

Введение

В последнее десятилетие в медицинской практике все более широкое применение находят различные электронные справочные системы, системы интеллектуальной поддержки деятельности врача при решении задач дифференциальной диагностики, прогнозирования и выбора плана и тактики лечения. При наличии разнообразного арсенала лекарственных средств и разработанных показаний к их применению практически очень трудно врачу выбрать те препараты, которые будут наиболее полезны больному в конкретной ситуации. Этим качеством владеют лишь опытные врачи с большим стажем практической аналитической работы. Поэтому использование новых информационных технологий для создания медицинских советующих систем, объединяющих знания и опыт врачей-экспертов, является важной задачей.

Проблема формирования базы знаний, включающей различные алгоритмы и модели диагностики, прогнозирования и принятия оптимальных решений при выборе тактики лечения, занимает центральное место при разработке компьютерных систем интеллектуальной поддержки деятельности врача. В связи с этим требуется разработка различных подходов, позволяющих повысить качество и надежность вычислительных процедур. Это возможно при использовании единой методики структуризации архивной информации, проведения системного анализа полученных данных, построения прогностических и оптимизационных моделей.

1 Формализованная постановка задачи

Одним из важных этапов лечения больных является этап подбора индивидуального медикаментозного лечения [1]. Он осуществляется врачом в течение времени $t = \overline{1,T}$ после проведения обследования пациента, включающего:

- сбора анамнеза;
- объективных исследований;
- лабораторных и инструментальных исследований;
- специальных исследований.

Эти данные используются для первичного принятия следующей последовательности врачебных решений:

- выбора групп лекарственных препаратов, необходимых для медикаментозного лечения;

- выбора определенного препарата в каждой группе;
- назначения суточной дозы лекарства;
- назначения контрольных точек в t^k день подбора индивидуального лечения для оценки его результативности.

Перечисленные решения принимаются на основе знаний, умений, опыта врача. Недостаточная результативность первичного назначения лекарственных препаратов корректируется в последующие дни подбора лечения с оценкой его результативности. Если результативность в очередной контрольной точке устраивает врача, то этап подбора индивидуального медикаментозного лечения завершается. Сократить число контрольных точек и создать предпосылки для подбора более адекватного состоянию больного варианта лечения позволяют методы интеллектуальной поддержки врачебных решений [2].

С целью применения этих методов рассмотрим формализованное описание изменения состояния больного в период $t = \overline{1,T}$ подбора лекарственных средств в виде следующей прогностической модели:

$$f_i(t) = \varphi_i\left(x, y, d_{mn}(t^1), d_{mn}(t^k), f_i(t^k), t\right) \quad (1)$$

где $f_i(t) -$ прогнозные значения показателей, характеризующих результативность лечения в $t -$ й день реализации этапа подбора, $i = \overline{1,I}$;

$f_i(t^k) -$ значения показателей в дни контрольных точек t^k;

$x = \left(x_1, \dots, x_{j1}, \dots, x_j\right) -$ вектор параметров, выявленных при сборе анамнеза и проведения объективных исследований;

$y = (y_1, \dots, y_l, \dots, y_L) -$ вектор параметров лабораторных, инструментальных и специальных исследований;

$d_{mn}(t^1)$ –значения суточной дозы n_m –го лекарственного препарата, $n_m = \overline{1, N_m}$ из m –группы, $m = \overline{1, M}$ при первом назначении;

$d_{mn}(t^k) -$ скорректированные значения суточной дозы в контрольных точках;

$\varphi_i(\cdot)$ –прогностическая функция.

2 Дуальномодельный подход к прогнозированию и оптимизации медикаментозного лечения

Предлагается в качестве основы интеллектуальной поддержки врачебных решений дуальномодельный подход к прогнозированию и оптимизации медикаментозного лечения. Дуальность определяется двумя способами использования формализованного описания (1):

- врач ориентируется на значения $f_i(t)$ для замены реальных значений $f_i(t^k)$ прогностическими оценками, что позволяет с меньшей частотой корректировать назначения лекарственных препаратов и тем самым сократить число контрольных точек;
- осуществляется трансформация (1) в модель, позволяющую включить врача в автоматический поиск лечения, и тем самым создать предпосылки для выбора более адекватного состоянию больного назначения лекарственных средств в режиме экспертного оценивания.

Рассмотрим реализацию первого способа дуального подхода. В этом случае, как показано в [2] эффективной является нейросетевая модель, определяющая зависимость (1) для первого назначения врача

$$f_i(t) = \varphi_i(x, y, d_{mn}(t^1), t) \quad (2)$$

С этой целью в строках обучающей выборки не включается ретроспективная информация $(d_{mn}t^k), f_i(t^k)$, а характеризация выбора врачом $n-$го лекарственного препарата из $m-$й группы достигается включением нулевых значений $d_{mn}(t^1)$, не использованных в индивидуальном назначении врача. На основе указанной обучающей выборки в системе STATISTICA формируется нейросетевая модель, отражающая зависимость (2).

Для подбора лечения нового больного с параметрами x^0, y^0 врача вносит эти данные в систему STATISTICA, проводит назначение $d_{mn}(t^1)$ и заполняет соответствующие столбцы. После этого появляется возможность вычисления прогнозных оценок результативности $f_i(t)$ для дней лечения $t^1 < t \le t^{k1}$, где $t^{k1} -$ день первой контрольной точки. Если врач удовлетворен прогнозируемой динамикой лечения на этом интервале времени, то он может не проводить коррекцию, а перейти к прогнозированию на интервале $t^{k1} < t \le t^{k2}$, где $t^{k2} -$ день второй контрольной точки по нейросетевой модели, отражающей зависимость (2). В противном случае он корректирует лечение и далее для прогнозирования использует нейросетевую модель, отражающую зависимость

$$f_i(t) == \varphi_i(x, y, d_{mn}(t^1), d_{mn}(t^{k1}), f_i(t^{k1})t) \tag{3}$$

Аналогичным образом осуществляется прогнозирование для последующих интервалов между контрольными точками.

Реализация второго способа дуальномодельного подхода основана на проведение в системе STATISTICA, модели (2), (3), активного имитационного эксперимента, позволяющего трансформировать эти зависимости в некоторую квадратичную модель [3].

$$f_i = \Psi_i\big(d_{mn}(t)\big), \tag{4}$$

где t принимают значения $t^1, t^{k1}, t^{k2}, \ldots$ В этом случае предлагается осуществлять поиск варианта медикаментозного лечения на основе многошаговой процедуры диалога с врачом и вычисления вероятностей значимости показателей результативности лечения на принятие решений при подборе лечения $-p_i$ [4].

На первом шаге принимается равная значимость, что соответствует значениям

$$p_i^1 = \frac{1}{I}\ \forall_i = \overline{1, I} \tag{5}$$

В соответствии с равномерным распределением значимости (5) случайным образом выбирается показатель i^1, для которого на основе зависимости (4) определяется оптимальный вариант лечения d_{mn}^{i1} и вычисляются значения всех показателей

$$f_i = \Psi_i(d_{mn}^{i1}), i = \overline{1, I} \tag{6}$$

Врачу задается первый вопрос: «Какое из значений показателей результативности лечения d_{mn}^{i1} необходимо изменить в первую очередь?». Пусть ответ - i_k, который формализуется следующей знаковой оценкой

$$A_i^k = \begin{cases} 1, \text{если } i = i_k, \\ -1, \text{в противном случае.} \end{cases}$$

Задается второй вопрос: «Увеличить или уменьшить?»; третий: «Сильно, существенно, несколько, немного, мало?»; четвертый: «Какое значение $f_{ik}^{\text{вр}}$ предлагается врачом?». Ответы на эти вопросы позволяют вычислить значение μ_i^k функции принадлежности лингвистической переменной $<$ необходимо изменить $>$ с термами, указанными во втором и третьем вопросах [5].

Полученная в диалоге с врачом экспертная информация позволяет скорректировать известное на $k-$ м шаге распределение

$$p_i^k, i=\overline{1,I}\ (k+1)-\text{го шага [6]:}$$

$$p_i^{k+1}=\frac{p_i^k+\varepsilon^{k+1}æ\left(A_i^k\right)}{1+\varepsilon^{k+1}}, i=\overline{1,I}, \tag{7}$$

$$\varepsilon^{k+1}=\varepsilon^k exp\left[\frac{\mu_i^k}{k} sign\left(A_i^k A_i^{k-1}\right)\right],$$

где $æ(A)=\begin{cases}1, \text{если } A>0,\\ 0, \text{если } A<0\end{cases}$;

ε^1- задается на первом шаге.

На каждом $k-$ом шаге повторяется случайный выбор по распределению p_i^k результативности, для которого решается задача оптимизации и осуществляется последующий диалог с врачом. За некоторое число шага K распределение значимости показателей (7) стабилизируются. Тогда значения вероятностей

$$p_i^K, i=\overline{1,I}$$

представляют собой некоторые весовые коэффициенты значимости

$$\alpha_i=p_i^K, 0\le\alpha_i\le 1, \sum_{i=1}^{I}\alpha_i=1,$$

на основе которых формируется интегральная оценка [6] результативности лечения

$$F(d_{mn})=\sum_{i=1}^{I}\alpha_i\,\widehat{\Psi}_i\,(d_{mn}), \tag{8}$$

где $\hat{\psi}_i(d_{mn})=\frac{\psi_i(d_{mn})-\psi_i^{\text{мин}}}{\psi_i^{\text{макс}}-\psi_i^{\text{мин}}}$,

$\psi_i^{\text{мин}}, \psi_i^{\text{макс}}-$ минимально и максимально допустимое значение $i-$го показателя результативности лечения.

Функция $F(d_{mn})$ является аддитивной комбинацией функций $\psi_i(d_{mn})$ и тоже является квадратичной, позволяющей определить оптимальные дозы лекарственных воздействий d_{mn}^* [3] для каждой контрольной точки оценки результативности

Выводы

Таким образом у врача появляется дополнительная информация, позволяющая принимать более эффективные решения, что и определяет роль интеллектуальной поддержки при подборе медикаментозного лечения на основе дуальномодельного подхода.

Литература

[1] Фролов В.Н. Выбор тактики лечения с применением математических методов / В.Н. Фролов. – Воронеж: изд-во ВГУ, 1977. – 120 с.

[2] Гафанович Е.Я. Интеллектуализация диагностики и лечения артериальной гипертензии на основе дуального экспертно-игрового оценивания / Е.Я. Гафанович. – Воронеж: НПЦ «Научная книга», 2011. – 112с.

[3] Гафанович Е.Я. Повышение эффективности экспертно-виртуального режима принятия врачебных решений на основе дуального моделирования / Е.Я. Гафанович // Системный анализ и управление в биомедицинских системах. – Т.16. - №3. – 2017. – С. 626-631.

[4] Львович Я.Е. Принятие решений в экспертно-виртуальной среде / Я.Е. Львович, И.Я. Львович. – Воронеж: НПЦ «Научная книга», 2010. – 140 с.

[5] Мельцер М.И. Диалоговое управление производством / М.И. Мельцер. – М.: Финансы и статистика, 1986. – 234 с.

[6] Львович И.Я. Информационные технологии моделирования и оптимизации: краткая теория и приложения / И.Я. Львович, Я.Е. Львович, В.Н. Фролов. – Воронеж: НПЦ «Научная книга», 2016. – 444 с.

Елена Гафанович
Candidate of медицинских Sciences, Associate Professor,
Саратовский государственный медицинский университет
им. В.И. Разумовского, Саратов, Россия
E-mail: lvovicha@mail.ru

Артем Львович
student,
Университет имени Я.А. Коменского, Братислава, Словакия
E-mail: artemlv2000@mail.ru

APPLICATION OF INFORMATION TECHNOLOGIES FOR VISUAL EVALUATION OF LEMNA MINOR BIOASSAYS

Oto Haffner, Erik Kučera, Štefan Kozák, Barbora Urminská

Abstract:

This article deals with the problem of automatic evaluation of experiments with Lemna minor. One of the convent methods for evaluating experiments with Lemna minor is number counting of new grown leaves of the Lemna minor. This article deals with the possibility of implementing algorithms of recognition and evaluation of Lemna minor leaves using visual system implemented in the single-board computers. Result of this paper is theoretical proposal of system for automated counting of Lemna minor leaves number.

Keywords:

ecotoxicity testing, image processing, single-board, Microsoft Azure, visual system, segmentation, IDS uEye XS, Raspberry Pi 2, machine vision

ACM Computing Classification System:

Computer vision, Visual inspection, Image segmentation, Cloud based storage

Introduction

The possibilities for monitoring the environment are in continuous progress and it is necessary to develop more advanced wastewater treatment methods and better monitoring of toxicant contents in treated water. Because of these facts, ecotoxicological tests are considered to be relevant approach for evaluation of impact of treated wastewater to ecosystems. No wastewater treatment plant removes pollution perfectly, it just lowers the pollution to an acceptable level. It is very important to judge whether the level of ecotoxicity of treated water is low enough and what will be the effect on the water organisms [8,10].

One of the plants which are used to measure ecotoxicity is Lemna minor. During the ecotoxicity tests the inhibition or biostimulation of plant growth caused by tested samples is watched for 7 days. The aim of the test is the quantification of substance effects on vegetable growth using number of leaves and biomass. Average increase of number of leaves in reference sample must be 7 times for the legitimacy of test. Also the pH of water must not change more than 1.5 unit and the level of EC50 for standard substance – potassium chloride KCl must be between 5.5 to 10 g L^{-1} [7].

The classical biotests utilizing duckweed are based on dry plant biomass comparisons between the control and experimental samples. These are invasive methods and do not allow the continuation of the experiments, since the plants are harvested for biomass measurements. Another method is based on manual counting of the number of thalli (‘‘leaves’’), regardless of their size. A more reliable and more accurate method of defining the growth of biomass is comparison of the area of thalli in the control and experimental samples.

This is a non-invasive method allowing the continuation of the experiment. Computer image analysis methods provide objective support for data collection. [11].

Big scientific institutions which deal with topic of wastewater management have devices with partially automated process of evaluation of Lemna minor experiments. Among these devices there are leaf area meters and chlorophyll fluorescence imaging instruments. However, these devices are expensive and not available for smaller institutions or scientific teams. Therefore the number of leaves at the end of the exposure period has to be counted manually by a laboratory worker. This is a very time consuming procedure and can be loaded by human mistake. Moreover, the unhealthy leafs can vary also in size and color – dead tissues have white color (necrosis), yellow color of leafs is caused by chlorosis – but these parameters cannot be quantified by human eye and are often overlooked. The majority of bioassays are based in a visual observation approach, where the observer evaluates symptoms of the toxic impact of pollution on bioindicators.

Machine vision has become a key technology in the area of manufacturing and quality control due to the increasing quality demands of manufacturers and customers. Machine vision utilizes industrial image processing through the use of cameras mounted over production lines and cells in order to visually inspect products, read or direct products and guide robots in real time without operator intervention. Machine vision can be used not only in industrial environment but also in various laboratories to support evaluation or analysis tasks, which would be difficult or time consuming if they were done by humans [11].

Using the digital images for the evaluation of the test would allow to measure the total leaf area and simultaneously to determine an average color of the plant. This method is therefore more accurate and much less time consuming compared to the manual counting of leaves.

Our study aimed to proposed a low-cost approach to objective and quantified biomonitoring based on computer image analysis, which can be used by a researcher without the need to purchase expensive specialized hardware. The aim of this study is to propose basic solutions for bioassay analysis using computer vision (standard digital camera) and framework software, which can be used in the future used for test populations of the widespread freshwater macrophyte Lemna minor L., treated within a toxicity bioassay of non-ionic detergents and compare the conventionally method (manual by research worker) and our new proposed. The function proposed algorithm was verified on set of lemna minor images.

1 Lemna minor

Lemna minor or lesser duckweed (Angiospermae, Lemnaceae) is a freshwater plant occurring in most countries of the world, especially in the lowlands and foothill areas with stagnant and slow-flowing waters. Body of the plant is made up of relatively long root and oval shaped leaves of 2 to 5 mm which float on the water surface. Plant usually creates the colonies of two to five leaves. A single lesser duckweed plant can reproduce itself about every 3 days under ideal conditions in nutrient-rich waters [9]. Bioassay on Lemna minor is quite common and easy to perform. The endpoint of the test is growth inhibition after 7 days exposure to tested samples and the aim is to quantify the effects of substances on vegetation growth.

2 Computer vision methods and algorithms

Computer vision is a field that includes methods for acquiring, processing, analysing, and understanding images. Computer vision applications range from tasks such as industrial machine vision systems to research into artificial intelligence and computers or robots that can comprehend the world around them. [2]

Computer vision covers the core technology of automated image analysis which is used in many fields. Machine vision usually refers to a process of combining automated image analysis with other methods and technologies to provide automated inspection and robot guidance in industrial applications. [1]

Thresholding

Thresholding is one of the most used method for image segmentation. Let us consider image as function f. Thresholding is then transformation of input image to output image g using

$$g(x,y) = \begin{cases} 1 & for\ f(x,y) \geq T \\ 0 & for\ f(x,y) < T \end{cases} \quad (1)$$

There are some methods for automatic setting of the threshold.

Shape modelling

This method uses two degree function for histogram approximation. Between this function and histogram is minimized sum of squares. The solution is found using iteration [6]

$$T_{opt} = \min\left[\sum_{g=0}^{T}[b_1(T) - g]^2 + \sum_{g=T+1}^{L}[b_2(T) - g]^2\right] \quad (2)$$

$$b_1(T) = m_0(T)/P(T) \quad (3)$$

$$b_2(T) = m_p(T)/[1 - P(T)] \quad (4)$$

Local thresholding

In some cases, global thresholding give us unsatisfied results. Advantage of local thresholding is respect of different light conditions in different parts of image.

The threshold is set for each pixel based on its surrounding. The imaginal mask is moved over the whole image. [5]

Binary mathematical morphology

Next morphological transformation will be expressed using Minkowski formalism. [4]

Dilation

$$X \oplus B = \{p \in \varepsilon^2 : x + b, x \in X, b \epsilon B\} \quad (5)$$

Erosion

$$X \ominus B = \{p \in \varepsilon^2 : p + b \epsilon X\ for\ each\ b \epsilon B\} \quad (6)$$

Opening and closing

Using combination of this two projection, we get morphological operation opening and closing. [5]

Opening is defined as dilation after erosion.

$$X \circ B = (X \ominus B) \oplus B \tag{7}$$

Closing is defined as erosion after dilation.

$$X \bullet B = (X \oplus B) \ominus B \tag{8}$$

RGB color model

The RGB color model (

Fig.1) is an additive color model in which red, green and blue light are added together in various ways to reproduce a broad array of colors. The name of the model comes from the initials of the three additive primary colors, red, green and blue. A color in the RGB color model is described by indicating how much of each of the red, green, and blue is included. The color is expressed as an RGB triplet (r,g,b), each component of which can vary from zero to a defined maximum value. If all the components are at zero the result is black; if all are at maximum, the result is the brightest representable white. [3]

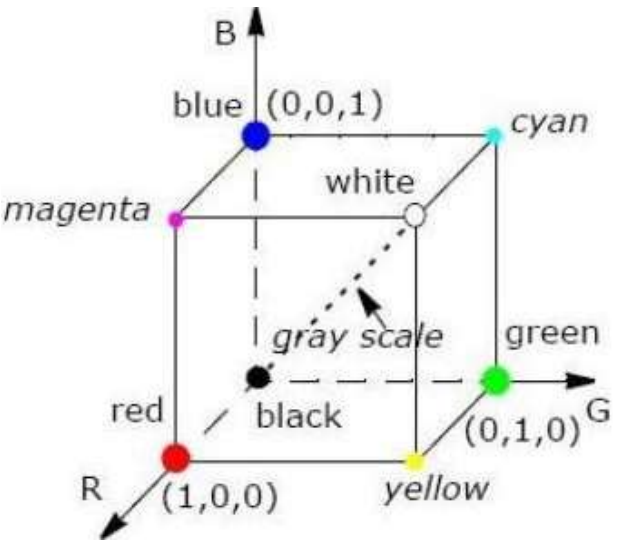


Fig.1. Representation of RGB color model

HSL/HSV color space

HSL and HSV color models (

Fig.2) are the two most common cylindrical-coordinate representations of points in an RGB color model. The two representations rearrange the geometry of RGB in an attempt to be more intuitive and perceptually relevant than the cartesian (cube) representation. Developed in the 1970s for computer graphics applications, HSL and HSV are used today in color pickers, in image editing software, and less commonly in image analysis and computer vision. HSL and HSV are the two most common cylindrical-coordinate representations of points in an RGB color model. The two representations rearrange the geometry of RGB in an attempt to be more intuitive and perceptually relevant than the cartesian (cube) representation. Developed in the 1970s for computer graphics applications, HSL and HSV are used today in color pickers, in image editing software, and less commonly in image analysis and computer vision. [6]

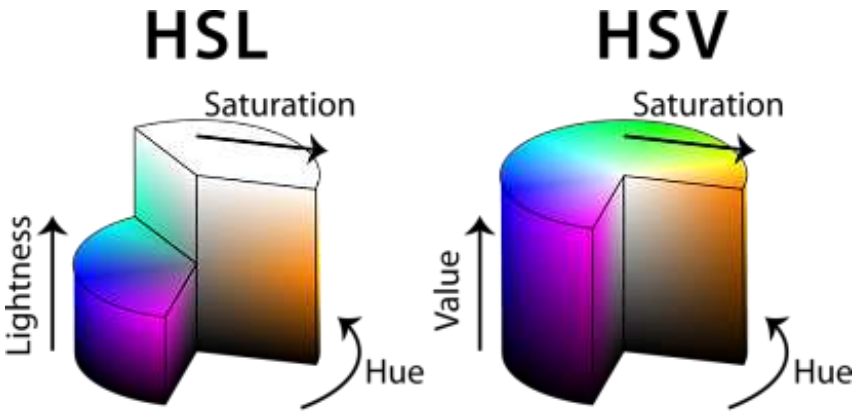


Fig.2. Representation of HSL/HSV color model

3 Proposal of methodology for experiment evaluation

The main motivation of computer vision Lemna minor evaluation is to reduce the mistakes caused by human. In some specific cases (e.g. small laboratories) the counting process of growth leaves is done by laboratory workers eyes. This can cause careless mistakes and is time-consuming.

Lemna minor specifications

The leaves are shown in Fig.3. In experiment, there are two containers. The first container is referral with clean water and the second is the same sample of contamined water with Lemna minor. In this experiment, the growing of number of leaves and the color of leaves is important.

Proposal of the method

For proposing of methodology, we will work with Lemna minor in laboratory environment.

Fig.3. Lemna minor laboratory experiment

Leaves segmentation

The first step is to segment the leaves in the image picture. The aim is to get only pixels in image which represent the area and shape of examined leaves (Fig. 4 left).

For segmentation can be used HSL color segmentation. This method is based on known color of object. The object can have various saturation of color e.g. Fig. 8 right, where the leaves are not so saturated and are faded.

In HSL color segmentation are very important set parameters for Hue, Saturation and Luminance. Not suitable parameters can cause unwanted results. In Fig.5 we can see segmented reflections in canister. To solve this problem by setting different values for HSL parameters can cause losing of information about leaves which have same HSL values as reflections. This problem will be discussed in further pats of paper.

In Fig.6 the results of wrong set Luminance parameter can be seen. The right parameters of HSL must be set experimentally and they also depend on lighting of the scene.

Counting of area

The better indicator than number of leaves is the area of leaves. Number of leaves are used because it is easy to count the number of leaves by laboratory workers. However, the leaves do not have the same size and this can cause some errors. The better way is compute the area of leaves. This task is very difficult for human laboratory worker, however for computer vision it is a very simple task. After watershed segmentation, the number of pixels are counted. The area of one pixel depends on used optics and camera sensor size. It is expected, that the level of water and leaves are still in the same distance from the camera. In this case, the simple measurement can be used to find out the real size of one pixel in image.

Leaves tint

One of the experiment indicator is color of leaves. In Fig.3 are shown two canisters. On the left side, there is non-toxic environment. On the right side, there is toxic environment. We can see, that color of leaves in toxic environment is desaturated and slightly yellowish. The exact color tint of leaves can be easily measured by meaning all HSL values of each pixel. For measuring the right color tint, it is necessary to have stable lighting environment.

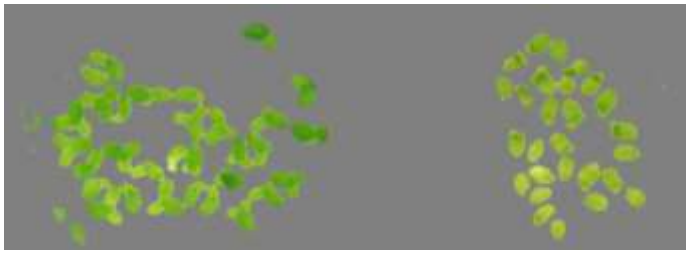


Fig.4. Lemna minor segmentation

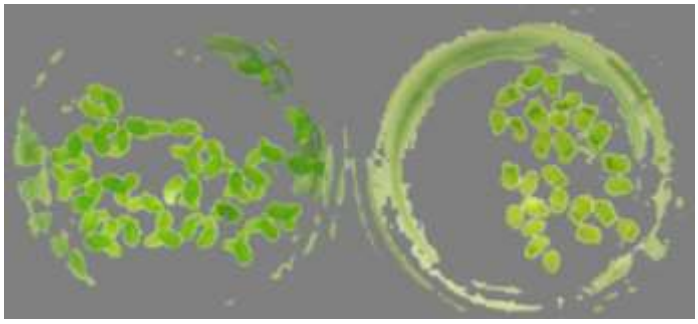


Fig.5. Lemna minor reflection segmentation

Fig.6. Lemna minor wrong set Luminance segmentation

Algorithm

In Fig.7 we can see the scheme of whole algorithm for experiment evaluation.

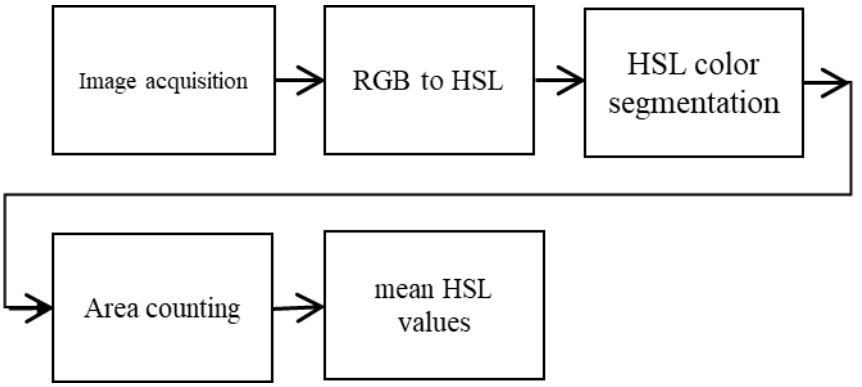


Fig.7. Lemna minor surface evaluate algorithm

4 Proposal of visual system experiment

Leaves tint

As we mentioned in section IV, there are a few problems in process of segmentation. The biggest problem are reflections in glass containers (Fig.8 left). These reflections cause mistaken segmentation, which cannot be removed by setting ideal HSL values. The possible solution is using not glass container but matt black. Container with mat black finishing would be ideal to suppress the reflections. Another solution can be using a region of interest in image. This can be used only when the position of camera and the container is still the same.

In Fig.9 we can see effects of water tint itself. Water from toxic environment can have color tint. This cause also error during segmentation. Container with mat black finishing would be ideal to suppress this error.

Luminance

In Fig.6 we showed the effect of non-uniform lightning to the segmentation. In case, that some parts of leaves are lighted non-uniform can cause error in segmentation. However,this can cause also errors in computation of color tint. This problem can be solved by proper lighting design.

In Fig.9. Water tint we can see effects of water tint itself. Water from toxic environment can have color tint. This cause also error during segmentation. These errors can be also removed with container with mat black finishing.

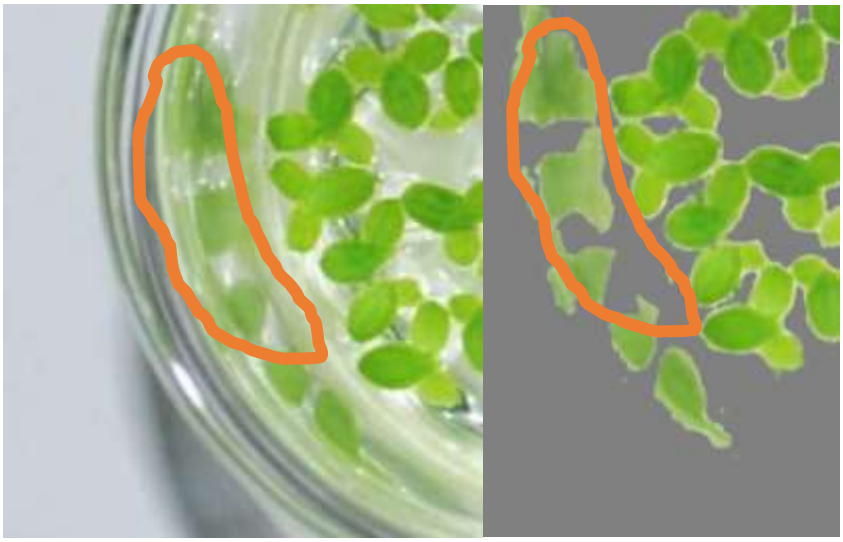


Fig.8. Reflections

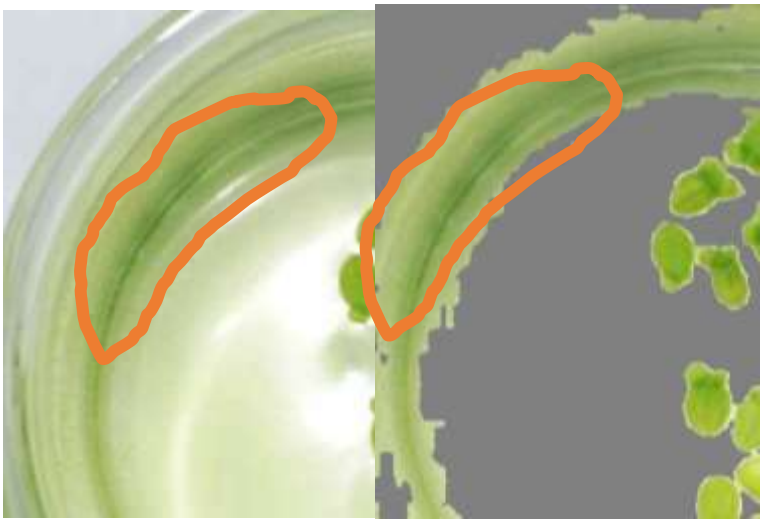


Fig.9. Water tint

Lightning of laboratory experiment

The lightning of the scene can be done in more ways. In

Fig.10 we can see the circle shape of lightning. Light with circle shape supplies more diffused light with direction from the camera lens. This configuration gives bright image field.

In Fig.11 there is a scheme with dome shape lightning. This configuration gives ale bright image field same as circle shape lightning. However, it gives much better diffused light. Adding of light beams inside of canonical surface makes shadowless light. The surface of *Lemna minor* leaves can be rough, so shadowless light can improve quality of scene image.

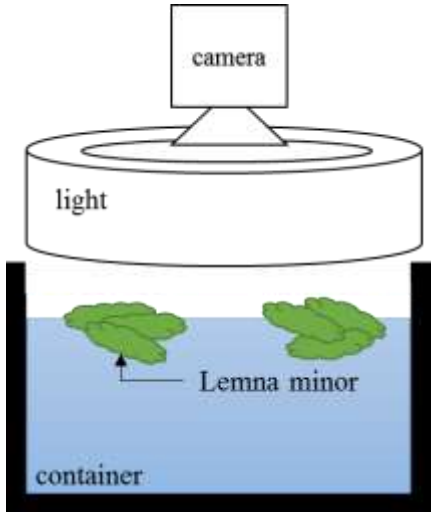


Fig.10. Circle lightning of laboratory experiment

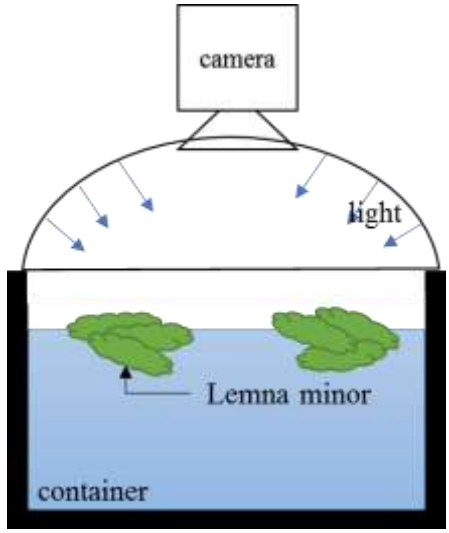


Fig.11. Dome shape lightning of laboratory experiment

5 Implementation

Proposed algorithm and visual system experiment must be implemented. The implementation scheme is shown in Fig.12

The whole system consists of single-board computer, visual system (camera) and cloud service. Camera is connected to the module for communication. The acquisitioned images are transformed from RGB color space to HSL. After transformation, the HSL color segmentation is done. Result of color regimentation is watershed segmented. After this, three computing tasks are done. To the .txt file is written the number of leaves, area of leaves and mean HSL value. To the .txt field it is also written value from pH probe. This .txt file is processed to be suitable for cloud communication module. This module sends the data to web service (database) to be accessible from any place or to be processed and evaluate in future.

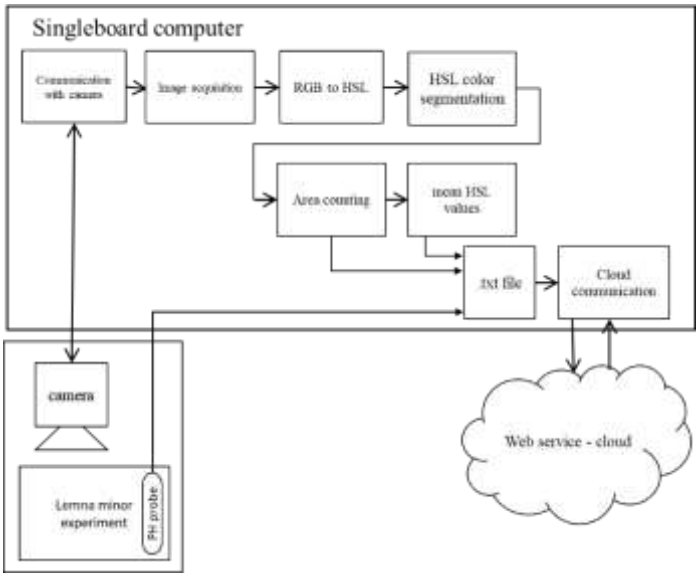


Fig.112. Dome shape lightning of laboratory experiment

Conclusion

In our work, we bring theoretical background about Lemna minor toxicity experiments. Based on this, the computer vison algorithms are presented, which can be used for morphometric measurements of plant surface. This approach can be new standard for a more objective and quantified assessment of the negative influence of toxins on bioindicators. Automation can increase the speed and accuracy of the tests evaluation. The next part of the work presents theoretical purposing of Lemna minor experiment examination and evaluation method.

Big scientific institutions which deal with topic of wastewater management have devices with partially automated process of evaluation of Lemna minor experiments. However, these devices are expensive and not available for smaller institutions or scientific teams. Our proposed algorithm would remove these deficiencies, because it uses low cost devices, can be implemented on any image processing framework (e.g. Matlab, NI Vision Assistant etc.) and will be easy to use. Using the area of leaves can be more precise than the number of leaves. Number of leaves does not take into account the fact, that the leaves can vary in size.

The future task of our work will be the implementation and evaluation of our algorithm comparing the conventionally method (manual by research worker) and our new proposed and its implementation to single-board computer.

Acknowledgement

This work has been supported by the Cultural and Educational Grant Agency of the Ministry of Education, Science, Research and Sport of the Slovak Republic, KEGA 030STU-4/2017, and by the Scientific Grant Agency of the Ministry of Education, Science, Research and Sport of the Slovak Republic under the grant VEGA 1/0819/17.

References

[1] E. Davies. *Computer and machine vision: theory, algorithms, practicalities.* 4th ed. Amsterdam : Academic Press, 2012. s. 871.

[2] O. Haffner. *Edge detection in image.* Bachelor thesis. Bratislava : FEI STU, 2011. s. 52 (in Slovak).

[3] R. Gonzales and R. Woods. *Digital Image Processing.* 3rd. Upper Saddle River : Pearson, 2008. s. 976.

[4] M. Klíma, M. Bernas, J. Hozman and P. Dvořák. *Image information processing*. Praha: Vydavatelství ČVUT, 1999. 177s (in Czech).

[5] E. Šikudová, et al. *Objects detection and recognition.* Praha : Wikina, 2011(in Slovak).

[6] M. Šonka,V. Hlaváč and R. Boyle. *Image processing, analysis, and machine vision.* Toronto : Thomson, 2008. s. 829.

[7] ISO 20079:2005. Water quality - Determination of the toxic effect of water constituents and wastewater on duckweed (Lemna minor) - Duckweed growth inhibition test.

[8] Mendonça, E., Picado, A.,Paixão, S., Silva, L., Barbora, M. and Cunha, M. A. (2011). The role of ecotoxicological evaluation in changing the environmental paradigm of wastewater treatment management. The 6th Dubrovnik Conference on Sustainable Development of Energy, Water and Environment Systems (SDEWES'2011). DOI: 10.13140/2.1.2168.3205.

[9] Free Floating Plants: Lemna minor L., lesser duckweed; Lemna trisulca L., startduckweed or ivy duckweed. Department of Ecology: State of Washington [online]. 2012. http://www.ecy.wa.gov/programs/wq/plants/plantid2/descriptions/lemmin.html

[10] Bundschuh, M. (2014). The Challenge: Chemical and ecotoxicological characterization of wastewater treatment plant effluents. Environmental Toxicology and Chemistry. Vol. 33, Issue 11, p. 2407.

[11] Mazur, R., Szoszkiewicz, K., Lewicki, P. and Bedla, D. (2016). The use of computer image analysis in a Lemna minor L. bioassay. Hydrobiologia.

Ing. Oto Haffner, PhD.
Faculty of Electrical Engineering and Information Technology, Slovak University of Technology in Bratislava, Bratislava, Slovakia
oto.haffner@stuba.sk

Ing. Erik Kučera, PhD.
Faculty of Electrical Engineering and Information Technology, Slovak University of Technology in Bratislava, Bratislava, Slovakia

Prof. Ing. Štefan Kozák, PhD.
Faculty of Informatics, Pan-European university, Bratislava, Slovakia

Ing. Barbora Urminská
Faculty of Chemical and Food Technology, Slovak University of Technology in Bratislava, Bratislava, Slovakia

International Journal
Information Technology Applications (ITA)
Volume 7, Number 1, July 2018

TEXT CATEGORISATION USING MACHINE LEARNING AND NAÏVE BAYES CLASSIFIER

Marek Špilda

Abstract:

The aim of this paper is to explore, implement and verify the functionality of Naïve Bayes classifier to classify natural language texts into categories. A sample classifier was developed to evaluate the results. For text cleaning, different text features were labelled and tests were performed with different combinations of cleaning methods to evaluate its effects on the categorization accuracy.

Keywords:

machine learning, text classification, Naive Bayes.

Introduction

With the rise of electronic media and the Internet, the number of email and chat interactions between companies and their customers has largely increased. Companies need to distribute a growing number of written customer requests to specific departments and their qualified employees. Categorization algorithms will make this process more efficient in distribution of inquiries to the right personnel.

Automatic classification can also be used in processing other various texts such as newspaper and magazine articles, detecting spam messages and various fraud schemas, categorize social media posts and many other types of written texts.

This paper aims to describe different types of machine learning methods and explain, how Naïve Bayes algorithm works and show experimental results of classification of movie reviews into two different classes (1 star and 10 star reviews) with different types of text cleaning methods including text-only, lower and mixed case text or using special labels for question and exclamation marks, numbers or emoticons.

1 Machine learning

Machine learning is a subset of artificial intelligence field, that contains algorithms able to learn without having to implement strict processing of data sets. Machine learning explores the possibilities to process large amounts of data, find hidden patterns between data features and use the learned knowledge to work with new and unknown data. The main machine learning methods are:

- supervised learning
- unsupervised learning
- semi-supervised learning
- reinforcement learning

Supervised learning uses examples with existing inputs and outputs to train and then predict the outcomes of new events.

Unsupervised learning may be used, when the outputs, categories or labels, are not known in advance. The algorithms look for hidden relationships and patterns in given data.

Semi-supervised learning is a combination between supervised and unsupervised learning. Usually the input consists of subset of data including related output results (categories or labels) as well as subset of data without its outputs.

Reinforcement learning requires feedback, using which the algorithm learns on the rightness of its outcomes and gradually iterates to more accurate results.

2 Naïve Bayes classifier

Naive Bayes classifier represents a statistical classifier and uses the supervised learning method. Classifier predicts the class by calculating each class probability based on the Bayes' conditional probability. Conditional probability is calculated using following equation:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{1}$$

A and B are some events. $P(A|B)$ is a probability of occurrence of event A, if B is valid. $P(B|A)$ is a probability of event B, if A is valid. $P(A)$ and $P(B)$ are probabilities of events A and B respectively. With the Naïve Bayes classifier, there will be k classes $C_1, C_2, \dots, C_k$. Each tested document will be represented as a set of words, $X = \{x_1, x_2, \dots, x_n\}$, where x_i are single words in the tested document. The classifier will predict, into which class C the document X belongs by calculating the probability for each class. The result will be the class with the highest probability.

Naïve Bayes uses the concept of so called "bag of words". This means the order of the words will not matter. Example givne, for the algorithm, a document "This is a good movie" is exactly the same as "This movie good is a". That is also the origin of the name "Naïve" – the probabilities are "naively" multiplied without considering the order of the words themselves.

To calculate the class, following equation is used, where c_{nb} is the predicted class and w_i is the word at position i. The positions are indexes of words in the tested document.

$$c_{NB} = \underset{c \in C}{\operatorname{argmax}} P(c) \prod_{i \in positions} P(w_i|c) \tag{2}$$

An issue may arise if a tested document contains words, which are not in the training set. This would mean working with zero probability, giving a zero final result. To overcome this issue Laplace smoothing is introduced, where the number of occurrences of each word is increased by 1.

With documents containing significant amount of words, many very small fractions will be multiplied, resulting in floating point underflow. Hence the calculation will be done in log space. The final equations will be as follows:

$$c_{NB} = \underset{c \in C}{\operatorname{argmax}} P(c) + \sum_{i \in positions} logP(w_i|c) \tag{3}$$

3. Sample calculation

To demonstrate how Naïve Bayes classification algorithm works, a simple experiment will be used. The goal of this experiment is to utilize Naive Bayes to classify movie reviews into two different classes (1 star and 10 star reviews).

The following dataset contains a total number of 5 simple documents: 3 belonging to the 10 star class and 2 to 1 star class.:

Class C_{10star} - 10 star class documents:

1. the movie was good.
2. I liked the movie, great actors
3. great movie, star actors

Class C_{1star} - 1 star class documents:

1. bad movie
2. bad performance, bad actors

First, an unordered set of all distinct words present in the training set is created – the bag of words defined in the previous part of this paper. From a data mining perspective, for each document in the training set a total number of occurences in the bag of word set is determined.

The bag of words representation of each of the classes is described in the next tables.

Table 1 – Bag of words representation of documents in class C_{10star}

#DOC	the	movie	was	good	I	liked	great	actors	star	bad	performance
1	1	1	1	1	0	0	0	0	0	0	0
2	1	1	0	0	1	1	1	1	0	0	0
3	0	1	0	0	0	0	1	1	1	0	0

Table 2 – Bag of words representation of documents in class C_{1star}

#DOC	the	movie	was	good	I	liked	great	actors	star	bad	performance
1	0	1	0	0	0	0	0	0	0	1	0
2	0	0	0	0	0	0	0	1	0	2	1

$$P(C_{10star}) = \frac{3}{5} \tag{4}$$

$$P(C_{1star}) = \frac{2}{5} \tag{5}$$

$$P(w_i\,|c) = \frac{n_i+1}{n+vocabulary} \tag{6}$$

where vocabulary represents the total number of distinct words in the bag of words and n represents the total number of words present in the class.

A sample document “very good movie” would yield following results for the two classes:

$$P(C_{10star}) = \frac{3}{5} * P(very|C_{10star}) * P(good|C_{10star}) * P(movie|C_{10star}) =$$
$$= \frac{3}{5} * \left[\frac{0+1}{14+11}\right] * \left[\frac{1+1}{14+11}\right] * \left[\frac{3+1}{14+11}\right] =$$
$$= 0.6 * 0.04 * 0.08 * 0.16 = \mathbf{0.0003072} \tag{7}$$

$$P(C_{1star}) = \frac{2}{5} * P(very|C_{1star}) * P(good|C_{1star}) * P(movie|C_{1star}) =$$
$$= \frac{2}{5} * \left[\frac{0+1}{6+11}\right] * \left[\frac{0+1}{6+11}\right] * \left[\frac{1+1}{6+11}\right] =$$
$$= 0.4 * 0.059 * 0.059 * 0.118 = \mathbf{0.0001643} \tag{8}$$

$$P(C_{10star}) > P(C_{1star}) \quad (9)$$

The calculated values of conditional probabilities are consequently compared and classified as belonging to the group with higher probality. Therefore the sample document „very good movie“ is labeled class C_{10star}.

4 Implementation and evaluation

Although there are many Naïve Bayes classifier libraries available, own implementation of classifier was programmed. The sample classifier for this project was programmed using Python programming language. Python was chosen for its simplicity in creating quick prototypes, ease of use and a steep learning curve.

Program creates class objects which hold the word strings (the bag of words) and related information for each of the classes. A separate instance of the class is also used for testing data, as the tested data are also treated as bag of words.

In order to be able to easily turn on or off each of the data cleaning methods, a strings utility was implemented. Using this, the program allows to test all possible combinations by looping through all the on/off switch combinations. The results for the whole run are then stored in one final table.

For result processing, python library Pandas was used. Pandas is a library to hold large data frames and allows fast data processing. In addition, one line csv and excel exports are within the available methods, which help to reduce the burden of non-project core related tasks.

The data set used was extracted from user reviews of randomly selected movies from the IMDB database. An experimental set containing 2999 1-star and 2999 10-star movie reviews was created. 2000 reviews from each class were used for training and 999 reviews from each class were used for testing.

In addition, several different data cleaning methods were used during data Preparation phase and all of their combinations were tested. Methods for converting to lowercase, extracting numbers to be treated separately, extracting positive and negative emoticons, extracting question and exclamation marks and also for removing all non-alphanumeric characters were implemented.

Another method to clean data is to remove stop-words. Stop words are words with no special significance to the information carried by the test. In English texts, words such as ‘a’, ‘the’, ‘it’, ‘is’, etc. are all considered to be stop-words. A stop-word removal was also used in some tests.

In the results table, only results subset with the most significant cleaning methods is shown.

Table 3 - result subset with cleaned testing data

Success	TotalTests	Total Match	tolower Case	only Letters	remove Stop Words
87,64%	1998	1751	FALSE	FALSE	FALSE
87,64%	1998	1751	TRUE	FALSE	FALSE
87,19%	1998	1742	TRUE	FALSE	TRUE
86,49%	1998	1728	FALSE	FALSE	TRUE
86,44%	1998	1727	FALSE	TRUE	FALSE
86,14%	1998	1721	TRUE	TRUE	FALSE
85,69%	1998	1712	TRUE	TRUE	TRUE
85,64%	1998	1711	FALSE	TRUE	TRUE

Interesting discovery was, that the algorithm showed event better performance, when only the training set was cleaned, while leaving the testing set untouched. The results of successful matches were up to 90%.

Table 4 - result subset without cleaned testing data

Success	Total Tests	Total Match	tolower Case	only Letters	remove Stop Words
90,49%	1998	1808	TRUE	FALSE	FALSE
89,84%	1998	1795	TRUE	FALSE	TRUE
88,29%	1998	1764	TRUE	TRUE	FALSE
87,64%	1998	1751	FALSE	FALSE	FALSE
86,54%	1998	1729	TRUE	TRUE	TRUE
86,09%	1998	1720	FALSE	FALSE	TRUE
85,79%	1998	1714	FALSE	TRUE	FALSE
82,48%	1998	1648	FALSE	TRUE	TRUE

As it turned out, the most significant cleaning method was conversion to lowercase. Leaving out non-alphanumeric characters had small impact, but not positive. Least significant cleaning methods were marking emoticons and marking question and exclamation marks and marking numbers. These are not presented in the results table, as the differences varied in no more than fractions of percentage.

The stop words removal was expected to slightly improve the process. Measurements show, this assumption is not true. Removing stop words in all cases resulted in slightly worse classification outcome.

In general, Naïve Bayes gives good results for classification. The algorithm is not difficult to implement and does not require significant processing power for classification. It shows to be a good and cheap method to be considered for natural language document classification tasks.

Conclusion

Even though the algorithm is relatively simple to implement, the results are satisfying. Since the effect of the labelling emoticons, question and exclamation marks and marking numbers, was insignificant, in commercial applications this would be of no meaning. Naïve Bayes should be used as its name suggests – naively calculate the probabilities without any extensive data optimization.

Natural language processing is expected extend the possibilities, how people interact with technology. In addition, it provides many opportunities for exploration and scientific as well as commercial experiments.

References

[1] Jurafsky, D., Martin, J. H. (Draft of August 28, 2017) Speech and Language Processing, Third edition draft, https://web.stanford.edu/~jurafsky/slp3/ed3book.pdf

[2] Busemann, S., Schmeier, S., Arens, R. G. Message Classification in the Call Center, DFKI GmbH, Germany

[3] www.python.org/doc

[4] Barber, D. (DRAFT November 6, 2012) Bayesian Reasoning and Machine Learning, Cambridge University Press, http://www.cs.ucl.ac.uk/staff/d.barber/brml/

Bc. Marek Špilda
Faculty of Informatics, Pan-European University in Bratislava, Slovakia
marek.spilda@gmail.com

International Journal
Information Technology Applications (ITA)
Volume 7, Number 1, July 2018

ACCESSIBILITY OF WEB PAGES

Lucia Vladovičová

Abstract:

In this article we are dealing with the usability and accessibility of websites in the administration of public or state institutions located in the territory of the Slovak Republic. Specifically, the article describes how the created application works to automate the web accessibility tests. Subsequently, the results of web accessibility testing are carried out by the created application, as well as the results of manually testing website accessibility, they are presented in our article. This web application is made with JavaScript and JQuery. The user interface is created using Hyper Text Markup Language and Cascading Style Sheets.

Keywords:

Accessibility of web pages, application, testing, feasibility

Introduction

We live in the age of information technology when almost every private or public institution has its website to be accessible to the general public. Despite the effort to create a page accessible to all user groups, institutions are many times unsuccessful in this area.

We often encounter web site developers they want to create sites that would be accessible to the general public and groups with some kind of disadvantage, but the problem is that they often fail to imagine about accessibility and usability of web pages in this area, they do not know clearly to identify what must a web site contain in terms of accessibility and usability, and what not.

Web designers are expected to have an awareness of usability and if their role involves creating markup then they are also expected to be up to date with web accessibility guidelines [1].

For the above reasons, we have decided to explore the issue of website accessibility and usability, and then we would like to suggest a way for website developers to test their sites in the future so that their designed and implemented sites meet website accessibility standards.

According to ISO 9241-11, usability is understood as the extent to which the product can be used with the highest possible efficiency and satisfaction, with way to achieve the desired user goals [2].

In Slovakia, the usability and accessibility rules of websites are summarized in rules called Web Accessibility Standards. These standards are based on the Web Content Accessibility Guidelines (WCAG), which were standardized by the W3C consortium.

WCAG bears a wide range of recommendations to make web content available to a broad circle of people with disabilities, including blindness and weak vision, deafness, learning disabilities, cognitive limitations, speech disorders. These recommendations deal with the accessibility of web content on desktops, laptops, tablets, and mobile devices. [3].

Standards created by the Ministry of Finance of the Slovak Republic are made up of several parties, which can be divided into three groups and these are:

- Page design rules - This set of rules discusses font contrast and formatting font
- Technical implementation rules - In this group, the standards are largely focused on the correctness of the display, and what should individual HTML tags contain, such as that each “img” element must contain alternate text that appears if the image is not displayed or this text is read by nearly-blind man with e-reader.
- Page Content rules - In this group, the document describes what must be presented in content accessibility on the web page, such as map of sites.

1 Application design

This application was designed to control website accessibility standards by law no. 312/2010 Z.z. the proceedings of the Ministry of Finance of the Slovak Republic, and namely according to Annex No. 1 to the Decree entitled as website accessibility standards.

The application was designed as a web application executable by any web browser supporting Javascript.

The application was implemented with JavaScript and JQuery library. The basic screen that appears when you start the application was programmed in HTML and CSS.

The application consists of a text window into which the user inserts an HTML code for checking and a window with a list of errors found in given HTML code. Below the input text box there are buttons for starting the HTML code validation. Using these buttons, the user can clear the error message, restore the page, or open the validator's guide in the new browser window.

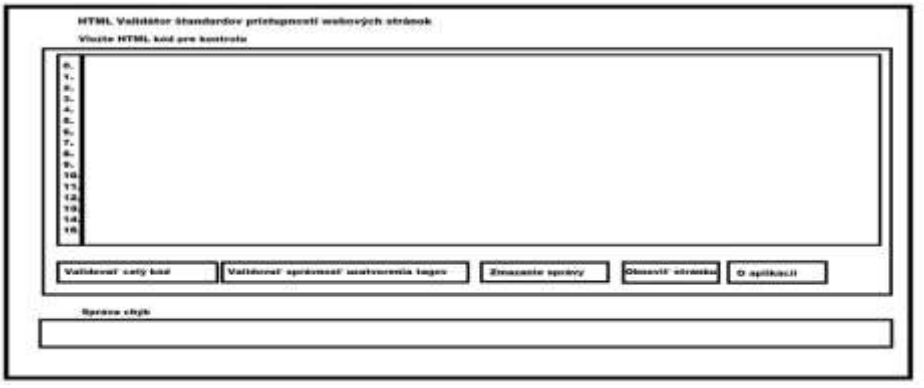


Fig.1. Final application design

2 Application functioning

The chapters below focus the principle of how an application works.

2.1 Validated embedded HTML code

The user opens the browser with validator and inserts the HTML code of the HTML page you want to test into a blank white window. The user has two options to check the embedded HTML code, to validate the entire code or only the correctness of HTML markup.

If a user chooses the button Validate the entire code, the validator retrieves the text entered by the user from the text area and checks the entire HTML code. This means that validator checks, whether the page does not contain prohibited HTML elements that, according to the website accessibility standards valid in the territory of the Slovak Republic, pages should not contain and also validator checks the allowed elements that have the prescribed structure. At the end, the selected feature will check whether HTML tags have correct close. Subsequent errors are listed in the box below the text area.

If the user just wants to check whether the HTML tags have correct close, user chooses the button to validate HTML tags correct close. When you start this feature, the app retrieves text from the text area to find if it is an HTML tag that is unpaired or if it is tag, which should have pair. In the case of a pairing tag, the first occurrence of the given HTML tag is loaded into the field and searches for the corresponding end tag. In the final step of this feature, validator detected errors on the incorrectly closed HTML tags, which are written by validator in error message with line number in code.

2.2 Line spacing in validator

For easier orientation in the embedded code, the app contains line spacing.

Inserting a line number into a text field is designed and implemented using the JQuery module, where we have created one more text area whose placement and display properties are defined using cascading styles via JavaScript.

The numbering method and algorithm is implemented on the principle that if it writes a line break in the text area to check the code, function adds one line and displays the line number in the next text area which is created next to the text area designed for HTML code.

2.3 Search HTML elements in code

To simplify the search for an error that was detected using the function "Validate the entire code" was implemented the HTML element search feature. This search function retrieves the whole text from the text area into the variable, in the next step divided the text into rows and searches for all elements with the given HTML tag and lists the line numbers where the tag was found. Subsequently, the user can find the line with error and fix the error.

2.4 Additional funcions of validator

The secondary small features include, for example, Deleting a message that the user can activate by clicking on the deleting a message button and use that functionality if user repeats validation of the code. Deleting message and repeat validation of the code, user find out what bugs have already fixed and what errors to fix, because the program does not automatically delete the old error message when re-checking, but current error message validator write under old error message.

Another secondary feature that a user can use by activating a button named Refresh the page is to delete all the data in all areas of the page in case they want to remove them. Before executing the action described, the user's validator alerts you to this fact and requires confirm user.

An indirectly related feature that a user can use in the event of a problem is feature under the About application button, where are the basic information on how to run JavaScript in your browser, the explanation of the individual button functions and the basic characteristic of the validator.

3 Analysis of selected pages

To make an automated testing using the application described above and then to make manual testing, we chose 5 webpages from either state or public institutions that should meet website accessibility standards, based on a document issued by the Ministry of Finance of the Slovak Republic in Standards for Public Information Systems reports, including a document on website accessibility standards.

3.1 Ministry of Finance of the Slovak Republic

After inserting the source code of the Ministry of Finance[1] site of the Slovak Republic into our validator, we found that three errors were found on the page in specific page elements. There was an alt attribute missing in one image and another attribute at the input button twice. When validating tags, the validator found 16 instances of incorrect closing tag.

If you are interested in eliminating the deficiencies and improving the accessibility of the site, we recommend that you correct the errors so that the page is correct or valid according to the W3C consortium, and we recommend that your optimize your site for mobile devices.

3.2 Comenius University in Bratislava

After inserting the Comenius University[2] site into our validator, we found shortcomings in the standard for providing alternative text in images of the so-called alt attribute, which should contain the description of the image. For twelve images, the alt attribute was not specified. Another drawback is that they do not have properly closed HTML tags by our validator.

We recommend removing shortcomings they have not correct close HTML tag and, since it is a college that specializes in the study of disabled students and wants to help them, it should add descriptions to each, which should contain alternative text in HTML elements as these descriptions have of great importance to the blind people.

3.3 Capital city Bratislava

After inserting the site of the capital city of Bratislava[3] into our validator, we found a mistake in alt attribute at elements input and several incorrectly closed tags. We recommend correcting all of these errors because this is a page on which is the assumption of attendance a large number of people with different devices and restrictions.

[1] The source code of the test page is located at: view-source:http://www.finance.gov.sk/
[2] The source code of the test page is located at: view-source:https://uniba.sk/
[3] The source code of the test page is located at: view-source:http://www.bratislava.sk/

3.4 City part Bratislava - Dúbravka

After inserting the city part Bratislava-Dúbravka[4] page into our validator, we have found incorrectly closed HTML tags, and the input element was missing a description in the form of an alternate text.

If you are interested in improving the usability of the site, we would recommend that all W3C error messages be removed, because incorrect closed HTML tag can have adverse consequences for correct display of the page on different devices.

3.5 Slovak television and radio

After inserting the Slovak television and radio[5] page into our validator, we found shortcomings in introducing an alternate text for the element "img" and the input element, namely, that the alternative text of these elements was not given. The site as one of the few have correctly closed HTML tags according to our validator.

In order to improve website accessibility, we recommend that you complement alternative text to HTML elements.

3.6 Evaluate results from manual site testing

For a better idea than the sites tested fulfilled the criteria set by the standards, we have compiled a summary table. In this table, we expressed the points that the test page fully meets or partially meets or does not meet the given criterion. If the web page has fully met the criterion test, the criterion obtained 1 point for this, if the page partially met the criterion, it gained 0.5 points if the page did not meet the criterion, it did not get any point.

Criteria/web page	**Ministry of Finance of the Slovak Republic**	**Comenius University in Bratislava**	**Capital city Bratislava**	**City part Bratislava-Dúbravka**	**Slovak television and radio**
Web accessibility Document	1	0	1	1	0
Map from sites	1	1	1	1	1
Alt attributes	0.5	0	1	1	1
Print version	1	1	0.5	1	0
Validate by W3C	0	0	1	0.5	0.5
Functionality without Flash and JavaScript	1	1	0.5	1	0
Correct structure table	1	0.5	0	0	1
Text version	0	0	1	1	0
Text accuracy in navigation	1	1	1	1	1
Optimization for mobile devices	0	1	0.5	1	0
Functionality without graphic	1	0	0.5	1	0.5
Forms	1	0	0	0	0
Documents in PDF, doc format	1	1	1	1	0
Total points	**9.5/13**	**6.5/13**	**10/13**	**10.5/13**	**5/13**

[4] The source code of the test page is located at: view-source:https://www.dubravka.sk/sk/Home.html

[5] The source code of the test page is located at: view-source:https://www.rtvs.sk/

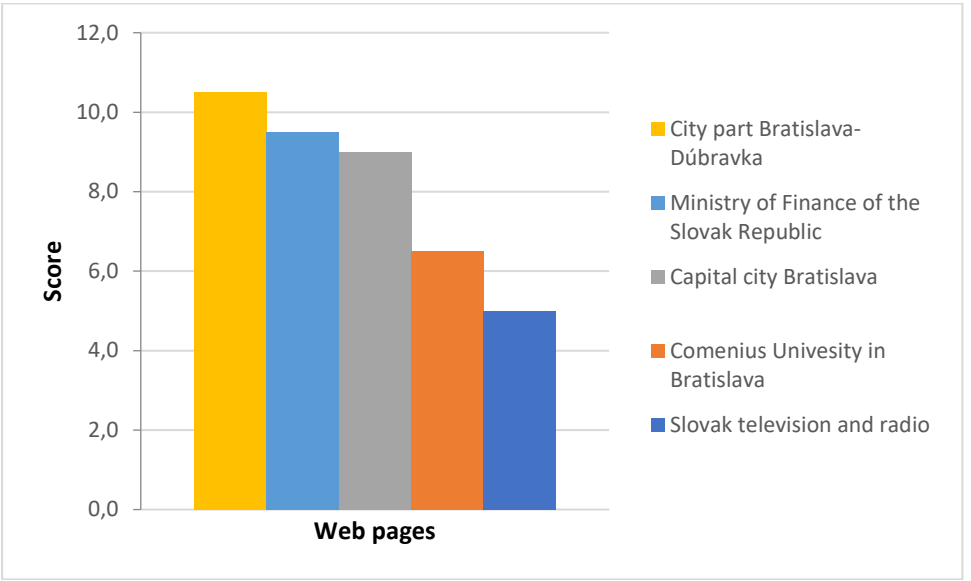


Fig.2. Graphic representation of the evaluation

The graphical representation shows that website accessibility standards from the tested pages were best met by the web page of City part Bratislava-Dúbravka, and the biggest gap in the fulfillment of the standards has the Slovak Television and Radio website.

Conclusion

The aim of this work was to design a possibility to create universal automated tests.

We have attempted to achieve this objective by studying the document entitled The Decree of the Ministry of Finance of the Slovak Republic on Standards for Public Administration Information Systems, which also contains web accessibility standards, and based on a detailed analysis of the document, we designed and subsequently implemented this proposal through the application.

We have tested the final version of the app as it works, whether there are any changes in display in the most commonly used web browsers across different platforms like Linux, Microsoft Windows and Apple. For Linux, we tested the functionality of the application in Mozilla Firefox or Opera. For Microsoft Windows, we tested the application in Google Chrome, Microsoft Internet Explorer and Microsoft Edge. For Apple platforms we tested the app in Safari.

In all tested browsers, the app appeared correctly as it should. All validator functions worked correctly, in addition to browser Mozilla Firefox, in which the main functionality of validator did not work, which ensures that the code is validated according to website accessibility standards, so we encourage browser Mozilla Firefox users to choose one of the alternative tested browsers to work with this app.

After testing the app in different browsers we looked at the application from user's perspective, so we found it convenient to write a manual. Based on this new look, we have written manual, which is part of the application called About App.

In the future, we are thinking of creating a site where validator will be created freely accessible to the public.

References

[1] Lacko Ján - Ružický Eugen. 2014. Web technologies and design. 1st ed. Praha: Wolters Kluwer 2014, 91 pages. [4,56 AH]- ISBN 9788074786327

[2] Makulová, S.2007. Použiteľnosť webových sietí ako základný predpoklad ich úspešnosti. V *Elet e-solutions* [online]. 2007.[cit. 2018-07-03]. Dostupné na internete: <https://www.elet.sk/blog/pouzitelnost-webovych-sidiel-ako-zakladny-predpoklad-ich-uspesnosti>

[3] Biro, P. 2009. Štandardy rýchlo a ľahko. V *informatizacia* [online].2009 .[cit. 2018-07-03]. Dostupné na internete: <http://www.informatizacia.sk/standardy-rychlo-a-lahko/>

Bc. Lucia Vladovičová
Faculty of Informatics, Pan-European University in Bratislava
lucia.vladovic@gmail.com

List of Reviewers

Issue 1/2018, in alphabetic order

Barabanov, V.F., Voronezh State Technical University, Voronezh, Russia
Beleckaya, S.J., Voronezh State Technical University, Voronezh, Russia
Ferko, A., Comenius University in Bratislava, Slovakia
Kashirina, I.L., Voronezh State Technical University, Voronezh, Russia
Kozák, Š., Pan-European University in Bratislava, Slovakia
Krajčík, V., College of Entrepreneurship and Law, Prague, Czech Republic
Lacko, J., Pan-European University in Bratislava, Slovakia
Ledeneva, T.M., Voronezh State Technical University, Voronezh, Russia
Novoselcev, V.I., Voronezh State Technical University, Voronezh, Russia
Ostapenko, A.G., Voronezh State Technical University, Voronezh, Russia
Pasternak, Y.G., Voronezh State Technical University, Voronezh, Russia
Rodionov, O.V., Voronezh State Technical University, Voronezh, Russia
Ružický, E., Pan-European University in Bratislava, Slovakia
Ryndin, A.A., Center of Information Technologies, Russia
Štefanovič, J., Pan-European University in Bratislava, Slovakia
Tolstykh, N.N., Concern "Sozvezdie"

Dipl. Ing. Michal Grell, CSc.
4.3. 1950 - 2.2. 2018

- for a long time the executive editor of our journal ITA have passed away this February.
Mr. Michal Grell was an expert for applications of mathematic methods in economy and in quantification of the impact of information technologies in market companies and public administration. He graduated at the School of Economics and Management of Public Administration in Bratislava. For a long time he was working as university teacher at the University of Economics in Bratislava, and as well at the school he graduated and at the Pan-European University in Bratislava too. He was an author and co-author of several scientific books, university textbooks and scientific works in domestic and foreign journals, where he contributed to the science and education. He was a founder and chairman of civil association EDUCATION-SCIENCE-RESEARCH in Bratislava, where he made a valuable contribution to the scientific-educational work.

ISSN: 2453-749X (online)

ISSN 1338-6468