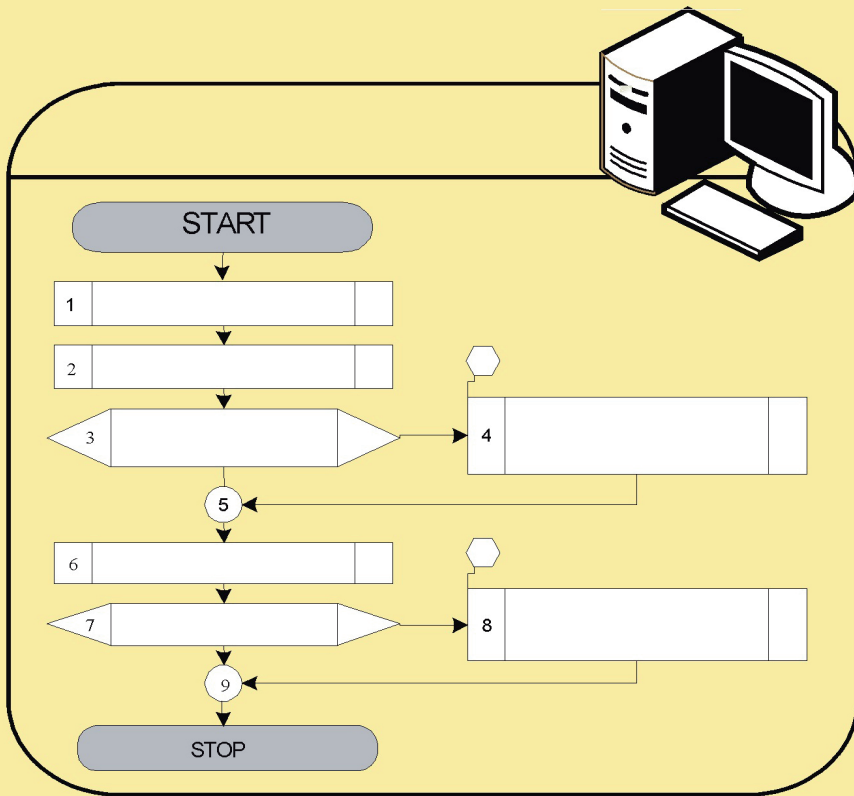


APPLIED COMPUTER SCIENCE

Vol. 11, No. 3, 2015



LUBLIN UNIVERSITY OF TECHNOLOGY

INSTITUTE OF TECHNOLOGICAL SYSTEMS OF INFORMATION

Scientific Board

Antoni Świć

Chair of Scientific Board

Lublin University of Technology, Poland

Zbigniew Banaszak

Deputy Chair of Scientific Board

Warsaw University of Technology, Poland

Józef Matuszek

Deputy Chair of Scientific Board

University of Bielsko-Biala, Poland

María del Carmen Alarcón del Amo

*Universidad de Castilla-La Mancha,
Albacete, Spain*

Josef Basl

University of Economics, Czech Republic

Grzegorz Bocewicz

Koszalin University of Technology, Poland

Krzysztof Bzdrya

Koszalin University of Technology, Poland

María Cordente Rodríguez

Universidad CEU - San Pablo, Spain

Milan Gregor

University of Zilina, Slovakia

Mikuláš Hajduk

Technical University of Kosice, Slovakia

Henning Heuer

Technische Universität Dresden, Germany

Petr Hořejší

University of West Bohemia, Czech Republic

Alexandr Jakimovič

Minsk State University of Technology, Byelorussia

José Mondéjar Jiménez

Universidad de Castilla-La Mancha, Cuenca, Spain

Martin Krajčovič

University of Zilina, Slovakia

Grzegorz Kłosowski

Lublin University of Technology, Poland

Emil Manoach

Bulgarian Academy of Science, Bulgaria

Egon Müller

Chemnitz University of Technology, Germany

Janusz Mleczo

University of Bielsko-Biala, Poland

Aleksander Moczała

University of Bielsko-Biala, Poland

Izabela Nielsen

Aalborg University, Denmark

Bogdan Palczewskij

Lutsk State University of Technology, Ukraine

Dariusz Plinta

University of Bielsko-Biala, Poland

Maria Reyes González

Universidad de Castilla-La Mancha, Albacete, Spain

Carlota Lorenzo Romero

Universidad de Castilla-La Mancha, Albacete, Spain

Ali Türkyilmaz

Fatih University, Turkey

Walter Terkaj

National Research Council

Editorial Board:	Arkadiusz Gola (Editor-in-Chief, Secretary of the Scientific Board)
	Jakub Szabelski (Proof Reading)
	Katarzyna Piotrowska (Editorial secretary)
	Dariusz Wołos (Editorial secretary)
	Lukasz Sobaszek (DTP specialist)
	Daniel Gąska (Webmaster)
Redactional Board:	Bożena Skołod - Computer Science in Production Engineering (Subject Redactor)
	Zbigniew Pastuszak - Computer Science in Economy and Management (Subject Redactor)
	Marek Janczarek - Computer Science in Technology (Subject Redactor)
	Alex Bodnar - English Instructor at Kyushu University, Japan (Language Redactor)
	Paweł Zaprawa (Statistical Redactor)
Publisher:	Institute of Technological Systems of Information, Lublin University of Technology
Editorial Office:	„Applied Computer Science” Editorial Office
	20-618 Lublin, ul. Nadbystrzycka 36, Poland
	Tel.: (+48 81) 538 44 83 Fax.: (+48 81) 538 46 81
	e-mail: acs@pollub.pl www.acs.pollub.pl
Circulation:	100 copies (the digital version is available at the journal's website: www.acs.pollub.pl and at Digital Library of Lublin University of Technology: www.bc.pollub.pl)

APPLIED COMPUTER SCIENCE

The Journal is a peer-reviewed, international, multidisciplinary journal covering a broad spectrum of topics of computer application in production engineering, technology, management and economy.

The main purpose of Applied Computer Science is to publish the results of cutting-edge research advancing the concepts, theories and implementation of novel solutions in computer technology. Papers presenting original research results related to applications of computer technology in production engineering, management, economy and technology are welcomed.

We welcome original papers written in English. The Journal also publishes technical briefs, discussions of previously published papers, book reviews, and editorials. Especially we welcome papers which deals with the problem of computer applications in such areas as:

- manufacturing,
- engineering,
- technology,
- designing,
- organization,
- management,
- economics,
- innovations,
- competitiveness,
- quality and costs.

The Journal is published quarterly and is indexed in: BazTech, CNKI Scholar (China National Knowledge Infrastructure), Index Copernicus, J-Gate, Google Scholar and TEMA Technik und Management.

Publication in Applied Computer Science is free of charge.

Letters to the Editor-in-Chief or Editorial Secretary are highly encouraged.

CONTENTS

Darina KUMIČÁKOVÁ, Alexander RENGEVIČ AUTOMATION OF MANUFACTURING TECHNOLOGIES WITH UTILISATION OF INDUSTRIAL ROBOTS	5
Patryk SOKOŁOWSKI, Jacek CZARNIGOWSKI, Paweł MAGRYTA CFD SIMULATION STUDY OF AIR FLOW AROUND THE AIRFOIL USING THE MAGNUS EFFECT.....	19
Irina KUSHNIRETSKA FORMING OF THE SEMISTRUCTURED DATA DYNAMIC INTEGRATION MASH-UP SYSTEM CONTENT.....	34
Rafał WOJSZCZYK THE MODEL AND FUNCTION OF QUALITY ASSESSMENT OF IMPLEMENTATION OF DESIGN PATTERNS.....	44
Natalya SHAKHOVSKA, Iryna SHVOROB THE METHOD FOR DETECTING PLAGIARISM IN A COLLECTION OF DOCUMENTS.....	56
Wojciech GRODZKI, Andrzej ŁUKASZEWICZ, Kacper LEŚNIEWSKI MODELLING OF UAV'S COMPOSITE STRUCTURES AND PREDICTION OF SAFETY FACTOR.....	67
Medjon HYSENAJ INTERACTIVE METHODOLOGY FOR SPATIAL REPLICATION AND RESEARCH.....	76
Iryna KARAYUZ, Petro BIDYUK FORECASTING GDP GROWTH RATE IN UKRAINE WITH ALTERNATIVE MODELS.....	88

Darina KUMIČÁKOVÁ*, Alexander RENGEVIČ**

AUTOMATION OF MANUFACTURING TECHNOLOGIES WITH UTILISATION OF INDUSTRIAL ROBOTS

Abstract

Industrial robots are important elements of automated production systems. Robots as programmable devices used in automatized batch or mass products production are often reprogramed to adapt themselves on the new product production changes. This time-consuming programming causes a decrease of manufacturing productivity. Therefore today robots can be reprogrammed by “offline” way too. The robot’s new control programme creation doesn’t limit the manufacturing process that is currently in progress at the plant. In addition, offline programming contributes to reduction of the period of the new product production starting. The article presents our practical experiences with utilisation of offline programme Fanuc Roboguide for solution of tasks of automation of welding, deburring and assembling technologies by industrial robots.

1. INTRODUCTION

Automation of the manufacturing process brings several benefits that contribute to improving its quality and productivity. Appropriate and well applied automation provides a solution for the product production costs decreasing and brings multiple gains. Utilisation of programmable devices allows to manufacturers to adapt on the new product production. However reprogramming of these devices takes the time to accomplish. This non-production time makes the products production in batches more expansive. Therefore today flexible production systems designed for variety products production require flexible reprogrammable devices for quick and automatic

*University of Žilina, Department of Automation and Production Systems, 010 01 Žilina, Slovak Republic, darina.kumicakova@fstroj.uniza.sk

** University of Žilina, Department of Automation and Production Systems, 010 01 Žilina, Slovak Republic, alexander.rengevic@fstroj.uniza.sk

change-over performing. CNC machines and industrial robots are such automatic devices. A flexible manufacturing systems (FMS) consists of several machines (CNC machine tools and/or industrial robots) linked to each other by a material-handling system. All aspects of the FMS are controlled by a central computer. CNC machine tools or industrial robots perform machining operations. But performing of inspection operations is also possible in automated inspection stations. A material-handling system, such as a conveyor system, or handling robots are capable to transport material (raw material, work pieces, parts, assembled product etc.) to the required position in the FMS. A central computer system is responsible for coordinating the activities of all automatically operating machines and the material-handling system in FMS. Although the FMS represents a high level of production automation, human labour is still needed to manage the system, load and unload parts, change tools, and maintain and repair the equipment.

In robotized manufacturing operations, the robot manipulates a tool (its end effector) to perform a manufacturing technology on the work part. The most widespread applications of robots for automation of manufacturing technologies can be found in the automotive industry. There a spot welding and spray painting are the most common applications of industrial robots. The other important manufacturing technologies automated by robots are: arc welding, grinding, polishing, deburring, pressworking, product assembly, etc. Robotic arm carries usually the special technological tool at its end and realizes this tool's movement on very complex paths to perform the required manufacturing technology. The robot programming performed directly at the automated technological workplace is the time-consuming process. This so called "online" programming causes downtime in production and thus contributes to reduction of productivity.

In conditions of batch production the more often reprogramming of CNC machines or robots are required. At the present the different programs of Computer Aided Process Planning (CAPE) are used to get the part program for CNC machine or control program for industrial robot in the shortest time and thus to prevent the production loss due to the occurrence of non-production times. Such programs are as follows [2, 7]: CAM (Computer Aided Manufacturing), CARC (Computer Aided Robot Control), CATS (Computer Aided Transport and Store), CAT (Computer Aided Testing) and CAA (Computer Aided Assembly). Currently the computer aid of manufacturing on NC machines is the most developed. Several worldwide producers offer either complex CAD/CAM systems for aid of a part designing and manufacturing or CAM systems specializing on the certain type manufacturing process (5-axis machining, machining by water or laser beams). These programs enable [1]:

- to create and simulate a technological procedure for machining the part of complex geometrical shape too,
- to transform the created procedure into CL data or to generate the postprocessor for translation CL data into the program for the CNC machine tool.

All these steps help to short the part machining time, to save material and to verify created technological procedure still before the part real machining. There we talk about offline programming of production machines. Computer aid of programming is also applied for industrial robots. Robot offline programing enables to create program for robot outside the production process on an external computer. Nowadays offline programs and their simulation tools help to create the optimal program paths for the robot to perform a specific task. Worldwide producers of robots offer usually offline programs only for own robot types.

2. ROBOT SIMULATION AND OFFLINE PROGRAMMING

Computer modelling and simulation is a very useful tool applicable in the process of designing, creation and evaluation of designed complex systems. Designers, program managers, or engineers can use computer modelling and simulation to understand and evaluate ‘what if’ case scenarios for models of a real or proposed system using computer software. Offline programming means programming robots outside the production – without stopping the production. It has grown the utilization rate of the robots in a short-run production companies. Finland Company Delfoi Robotics claims that industrial companies manufacturing with Delfoi software have this growth from 30% to over 90% [3]. Software for offline programming exploit data of 3D CAD models for the faster generation of programs. Due to calibration features the created programs can be so precise that the production of new products starts very quickly. The shortening of the time for adoption of new programs from weeks to a single day enables the robotization of short-run production.

Generally, offline robot simulation software is built on the Virtual Robot Controller (Virtual Robot Technology) which is the exact copy of the real robot’s controller. Therefore robot program and configuration parameters can be easily transferred between PC and the robot’s real world. Computer model of virtual robot usually consists of three components (Fig. 1) [5]:

- model of the robot’s arm (1),
- model of the control unit – virtual Teach Pendant (2) and
- program for performing the robot task (3).

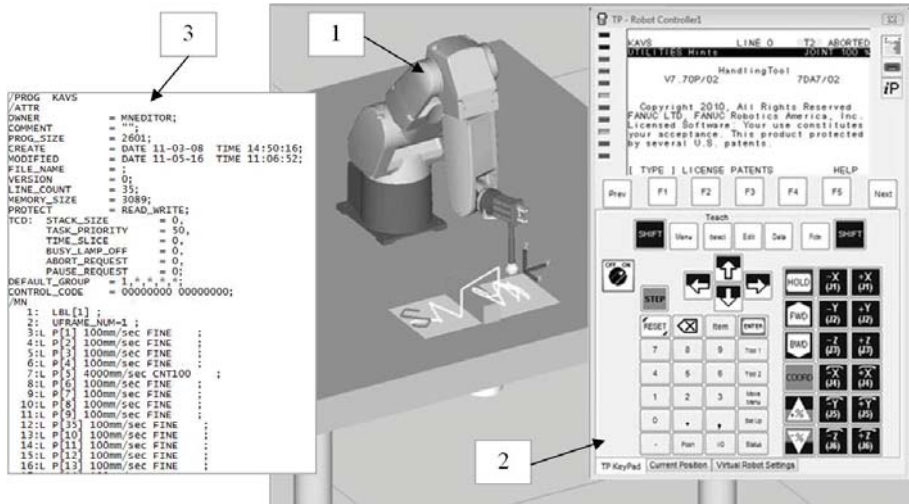


Fig. 1. The virtual robot computer model [8]

The software's simulation and control features enable a validation of the robot program still in preproduction stage and so to eliminate undesirable collisions, to optimise the robot's working cycle and to verify the robot's ability to reach all required positions in the work cell.

Robot offline programming software coming from robot manufacturers include RobotStudio (ABB), KUKA.Sim Pro (KUKA), MotoSIM EG-VRC (Yaskawa), RoboGuide (FANUC), etc. [10]: RobotStudio enables robot programs to be prepared in advance, increasing overall productivity. KUKA.Sim Pro is developed for offline programming of KUKA robots. It uses the build-in tools to load CAD data from other systems into KUKA.Sim Pro or build models using the system's CAD tools. The grippers, welding guns and other kinematic structures are possible to build too. Virtual KUKA controller allows the cycle time analysis. MotoSim EG-VRC (Motoman Simulator Enhanced Graphics – Virtual Robot Control) is a comprehensive software package that provides accurate 3D simulation of robot cells. This simulation software can be used to optimize robot and equipment placement, as well as to perform collision detection, reach modelling and cycle calculations. Fanuc Roboguide is a cost effective software package that enables to design, prove and program robot systems in real time, either concurrently in a new product design or for making modifications to existing equipment or programs. Roboguide can import unique CAD models of process parts and create a virtual work cell which includes machinery, part transfer devices and factory obstacles. Roboguide will then teach robot paths to simulate the operation and performance of the robotic application taking into account the physical obstacles. Reach verification, collision detection, accurate cycle time estimates and other visual system

operations are also simulated in Roboguide's 3D environment. These functions can be used to produce a simulation video to help evaluate and approve a project. Delfoi Robotic offline programming includes universal robot support and versatile and effective tool for a several areas in industrial robotics applications [3]: Delfoi ARC for ARC welding and laser welding, Delfoi CUT for cutting, milling and machining and Delfoi PAINT for painting and coating. Act/Weld ALMA (France) is offline program developed only for ARC welding application. It support of different types of robots (OTC Daihen, STAUBLI, ABB, KUKA, FANUC, PANASONIC etc.). The offline simulation software is exploited by small and medium-size enterprises and also by large world-class manufacturers.

3. AUTOMATION OF MANUFACTURING TECHNOLOGIES

The Department of Automation and Production Systems at the University of Žilina has several academic versions of offline programming software Fanuc Roboguide available. Roboguide is a robot system animation tool specifically developed for production and maintenance of robot systems. It can be used in offices or on the factory floor. Roboguide allows users to simulate the entire robotic cell or process in a virtual 3D space. The software is designed so that the robot and teach pendant operate exactly like a real robot. The simulation occurs by using a virtual robot. The standard software's modelling function was developed with the aim to reduce time for modelling devices. CAD data can be imported to create the parts by the modelling function. A large library allows the user to select and modify the parts and dimensions necessary. Roboguide incorporates many application-specific tools into its software options [8]:

- Roboguide HandlingPRO allows users to simulate a robotic process or study feasibility options for applications in a 3D space.
- Roboguide PaintPRO is a graphical offline programming tool for a path teaching and paint process development.
- Roboguide WeldPRO (SpotPRO) enables to simulate the robotic arc welding (spot welding) process in a 3D space.
- Roboguide PalletPRO enables the user to completely build, debug, and test palletizing or depalletizing application offline.

Department of Automation and Production Systems has available the module Roboguide HandlingPRO, which is a full-featured tool for creating automated technological workspaces [5]. This software option is most of all used in educational processes. But some problems of industrial practice were partially solved with support of program Roboguide HandlingPRO. The proposals of robotic cells for automation of arc welding, deburring and assembly technologies were solved.

3.1. Automation of arc welding technology by using a robot

The aim of the project was to propose solution to automate the workspace for welding steel construction using the method MAG (Metal Active Gas) [4]. The whole process of welding was carried out manually. The priority objective was to reduce the time of production associated with arc welding technology. Whole construction with dimensions 1725 x 600 – 700mm, consists of 54 components (Fig. 2), made of steel with guaranteed weldability to 25 mm of wall thickness. Butt and fillet welds were applied. For welding MIG/MAG method the ISO 6947 standard specifies seven welding positions: PA, PB, PC, PD, PE, PF and PG. For the welded construction three welding positions were used: PA, PB and PG (Fig. 3). Welding parameters (a size of welding current, a wire feed speed and a welding speed) were determined from templates of welding parameters for butt and fillet welds (Tab. 1). Based on this specification, the welding parameters were determined by total time required to weld the whole construction. Welding time t_w was calculated with utilisation of equation for the welding speed (v_w) calculation:

$$v_w = \frac{x_i}{t_w} \Rightarrow t_w = \frac{x_i}{v_z} \quad (1)$$

The calculated total welding time t_w was equal to 30 minutes and 58 seconds.

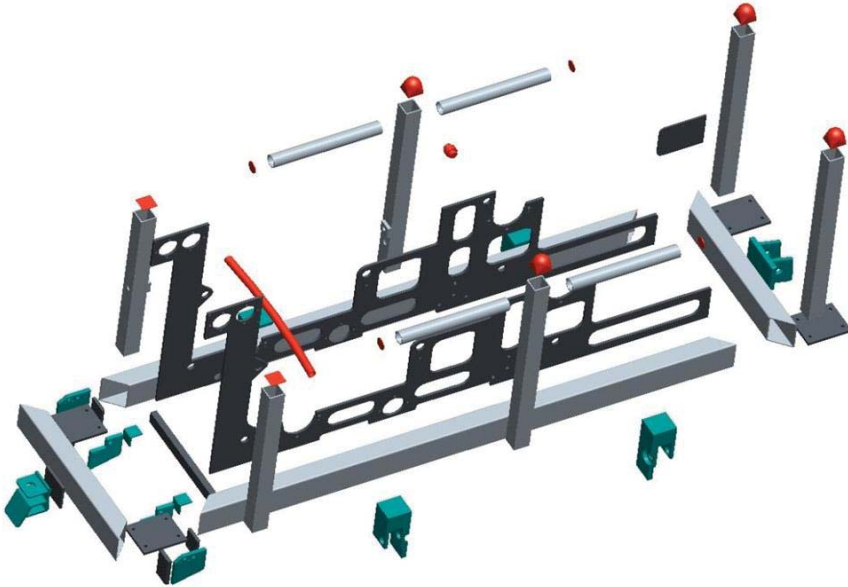


Fig. 2. Components of welded construction [4]

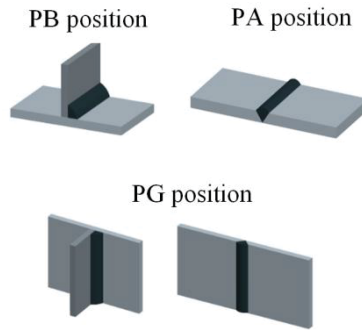


Fig. 3. Applied welding positions [4]

Tab. 1. Welding parameters [4]

i	Type of weld	Welding positions	Total length of welds x_i [mm]	Welding current I_w [A]	Wire feed [m.mm ⁻¹]	Welding speed [cm.min ⁻¹]
1		PB	2322	290	10,1	40
2		PG	60	230	7,3	28
3		PB	2240	290	10,1	40
4		PG	2560	230	7,3	28
5		PB	580	290	10,1	40
6		PB	1280	210	10,2	76
7		PA	800	120	4,4	27
8		PA	640	120	4,4	27
9		PG	800	150	6	46

During the selection of welding robot the following requirements were applied: a size of the robot arm reach with respect to the welded construction dimensions, robot arm mobility in six axes, routing of wire harness inside the robot arm and possibility to simulate a welding torch path in the simulation program Roboguide. For experimental purposes the robot Fanuc – type ACR Mate 120iC/10L with a size of the arm reach equal to 2009mm was selected [4]. For arc welding technology MAG type was selected with welding head FRONIUS Robacta 400, which meets the requirement of current load 290A. Accuracy of the position of the end point is $\pm 0,5\text{mm}$. Moreover, this type of welding head is directly designed for welding by using robots Fanuc, so that all welding parameters can be set directly from the Teach Pendant.

The production process of the welded construction is divided into two interlinked workplaces:

- the manually one, where worker performs all preparatory, auxiliary and technological works and
- the automatic one, where the robot performs a welding technology on the construction.

3D model layout of robotic welding cell was created with system Pro/Engineer, in which intended installation of welding robot on a portal construction – above the welded construction. To simulate the path of movement of the robot welding torch during the implementation of arc welding technology the Roboguide HandlingPRO was used. This program offers a virtual library containing models of welding robots. A simplified 3D model of workplace (model of welded construction, simplified model of the fixture for alignment and fixation construction skeleton) and welding head model were imported, in the format IGIS, into created workcell (Fig. 4).

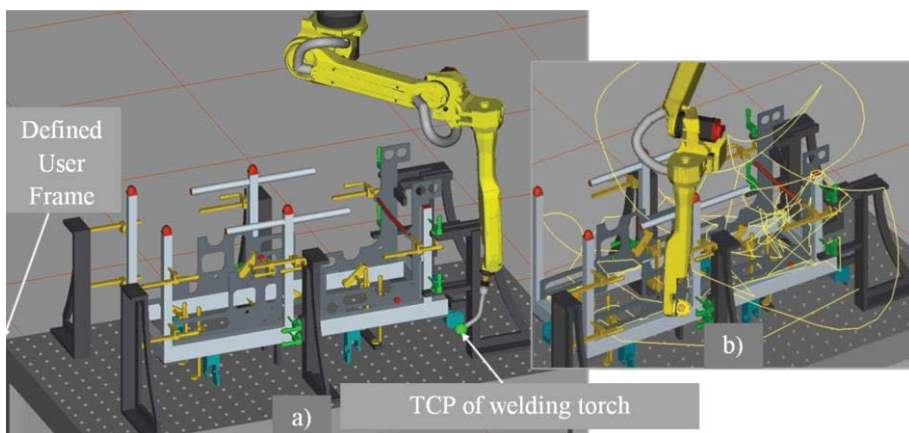


Fig. 4. Model of robotic workcell created in Roboguide Handling/PRO (a); TCP path movement displaying (b) [4]

HandlingPRO does not allow write all necessary welding parameters, but when a designed robot for arc welding is inserted, the INST (instruction) offer will be extended about possibility of turning on/ turning off welding current and voltage. At instruction “Arc End” the endurance on the end position of the weld before further movement of TCP (Tool Center Point) is written. When in the assembly the welds have identical parameters, size of current and voltage they can be entered in the register labelled ARC. During the simulation of welding robot operating cycle the reachability of all working positions with welding torch was tested. The duty cycle of the robot consists of two parts: stitching the skeleton of the construction and then its welding. Process simulation has shown that the time needed for stitching is $t_s = 3$ minutes and 30 seconds, and time for welding is $t_w = 20$ minutes and 48 seconds. The total operating cycle of robot usage for arc welding technology is 74% of t_w and remaining 26% is the time required to change position between welds.

3.2. Automation of deburring technology with robot

The task of automation of a finishing manufacturing technology used for deburring of a cast was solved for the purpose of preparation a laboratory training workplace at the Department of Automation and Production Systems. For representative workpiece has been proposed structural and layout solution of this workplace consisting of [9]:

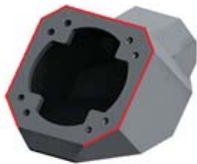
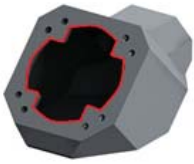
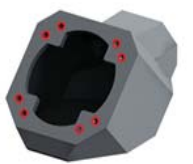


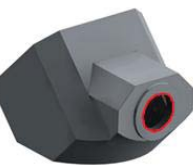
- 6 axis robot Fanuc LR Mate 200iC,
- system of automatic exchange of robot’s effectors Schunk,
- deburring head Flexdebur FDB 300 for radial deburring and
- modular system of clamping elements Siegmund.

3D model of workplace and the workpiece was created in system Pro/Engineer. The proposed workpiece is made from casted steel STN 42 2670/Z1. Considering the complexity of shape and method of workpiece clamping in fixture the technological process of the edge deburring by using a robot was carried out on two sides of the workpiece. Deburred edges and sequence of their machining are shown in Table 2 [9].

For the simulation purpose of the robot Fanuc LR Mate 200iC movement paths a workcell (Fig.5) in Fanuc Roboguide HandlingPRO was created. 3D models of components of designed robotic workstation created in system Pro/Engineer were imported in IGES format into the workcell. Virtual robot model was applied from the off-line program library. During robot programming in Roboguide HandlingPRO the function CAD-To-Path was used by which it is possible to define the coordinate points. By choosing Edge Line function the defining path of the tool along the edge of the workpiece is easier to define, because it automatically detects the edges of the workpiece. It is necessary to

select the first point on the edge and proceed in certain intervals as shown in figure 3. Then it creates a selection Feature in which necessary settings can be defined. The main function is to generate Teach Pendant program in which all settings are taken into account and according to these settings the control program is created. Thus all programs were created for deburring of defined edges on the workpiece. The tool path for automatic deburring can be seen in figure 6.

Tab. 2. Sequence of the workpiece's edges deburring [9]

Machining of the workpiece from the one side	1.	2.	3.
			
Machining of the workpiece from the other side	4.	5.	6.
			

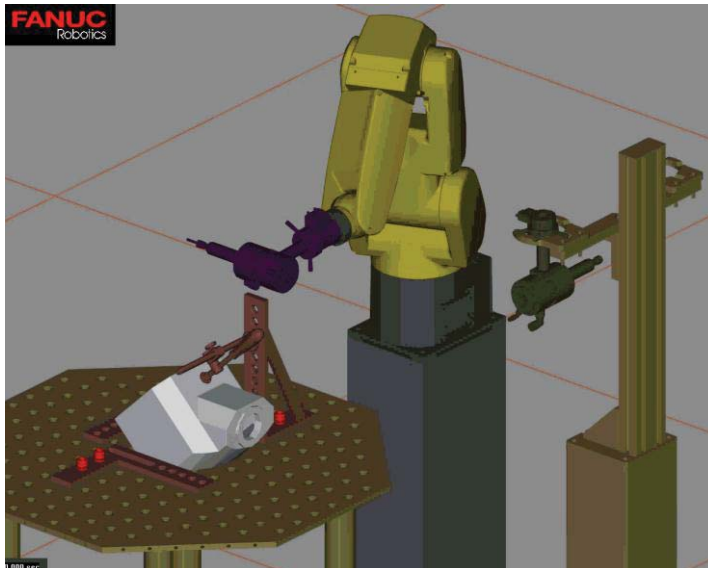


Fig. 5. Designed robotic cell for automation of deburring technology (Roboguide HandlingPRO) [9]

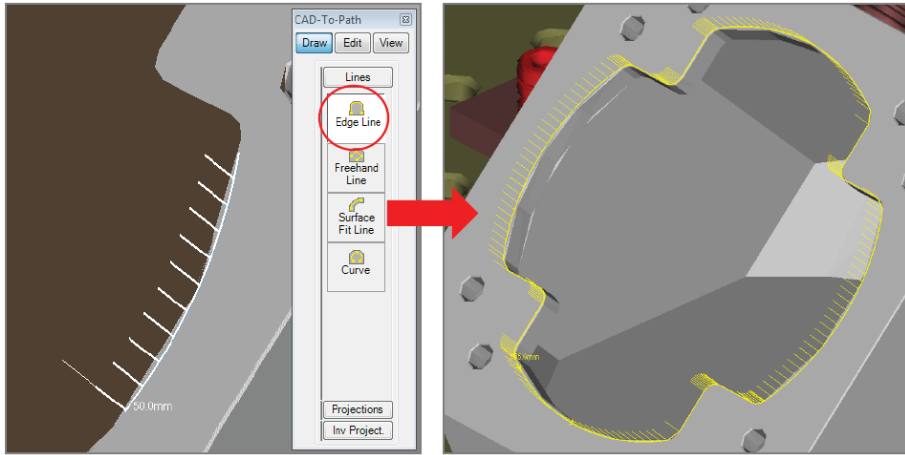


Fig. 6. Utilisation of *Edge Line* for the tool path definition [9]

Simulation of deburring tool path showed that robot is not able to deburre all edges on the workpiece as it was required. Fixation of the wokpiece in a fixture without possibility of its additional rotational movements does not allow the deburring tool to achieve edges at the bottom of the workpiece with respect to required technological conditions of deburring technology.

3.3. Automation of assembly tasks with robot

Automated assembly preparation constitutes a difficult tasks related with shape and dimension diversity of structural elements of assembled products [6]. The type of assembled elements (whether they are simple components or complex assembled subassemblies) and applied assembly methods influence the election of assembly and auxiliary equipment. Industrial robots are flexible assembling devices that are able to automatically preform main and auxiliary assembly operations. In field of automated assembly robots are applied most frequently for:

- Pick&Place applications (SCARA and Delta robots)
- Complex handling with assembled elements (6-axes robots) or
- Realisation methods of assembled elements joining (screwing, pressing, etc.)

At the Department of Automation and Production systems were built laboratory workstation of automated assembly for the purposes of teaching and research (Fig.7a). There the robot Fanuc LR Mate 200iC performs the final assembly operation of assembled product (Fig. 7b) and handling with assembled product. Robot takes it from the output of automated assembly workplace and storages it into the shelf storage. The final assembly operation is an operation of

screwing the cover with a diameter 40 mm. The finally assembled product is grasped in its parallel planar surfaces by specialised gripping jaws of the robot parallel gripper MHZ2-20D, SMC Industrial Automation. The grasped dimension is 45 mm.

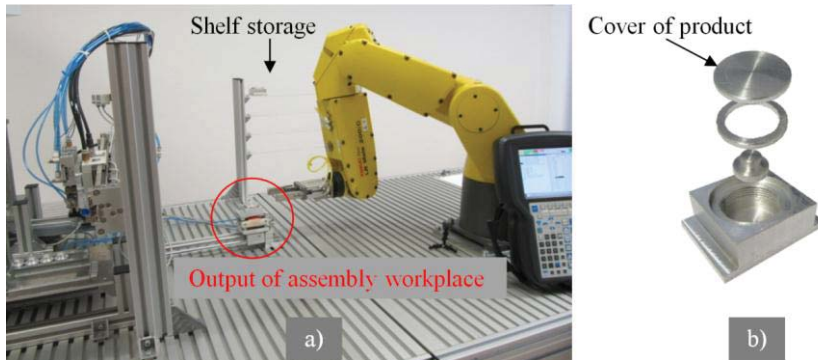


Fig. 7. Real Laboratory workstation of automated assembly (a); the real assembly product (b) [source: own study]

Creation of the robot Fanuc work cycle program was done in simulation program Fanuc Roboguide v.8 with HandlingPro option. 3D model of workplace (excluding the robot Fanuc model) created in CAD/CAM/CAE program Pro/Engineer Wildfire 4.0 was importing into Roboguide in IGIS. In simulation environment of HandlingPRO other additional objects were added into created Workcell to define them a motion by utilisation of function Machines. The virtual model of the robot Fanuc LR Mate 200iC has been implemented into Workcell from the offline program library. The procedure of control program creation has been divided into three separate programs: program of screwing of the cover, program of the assembled product taking out from the position of its final assembling and program of the assembled product storing into the shelf storage. All these programmes were formed the main control program. This approach to the robot control program creation gave us the possibility to edit the control program during its verification at the real laboratory workstation by much easier way. Operating of the shelf storage by the robot Fanuc is a sequential putting of final products on the single positions of each store shelves. Total storage capacity is 20 pieces of final product. Every shelf has 5 store positions. Offline program Roboguide offers special options for this operation type programming as follows: PalletPRO or PaletTool. We could work only with HandlingPRO. Therefore, we have applied the method of a gradual defining of the single store positions through the program's registers and position registers. The problems with Fanuc robot arm singularities during imposing the final products on lower positions of the shelves storage detected by simulation led to change of the shelves mounting configuration in the real shelves storage.

Figure 8 displays the workplace of robot Fanuc in conditions of HandlingPRO simulation environment (Fig. 8a) and the real robotized workplace (Fig. 8b).

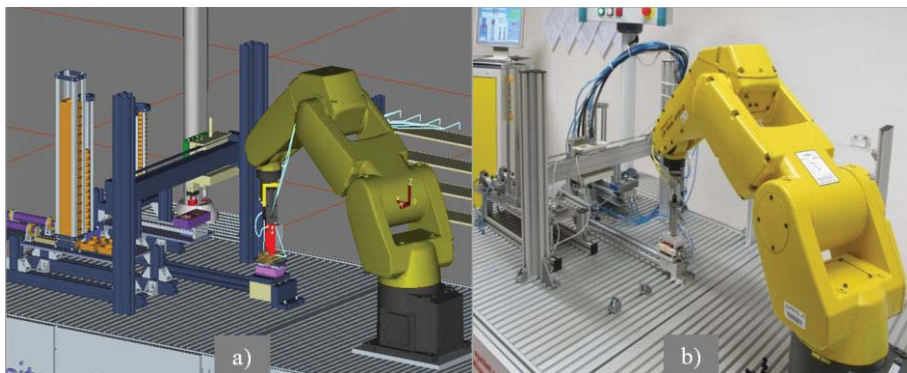


Fig. 8. Off-line robot programming in Roboguide HandlingPRO (a); Robot Fanuc LR Mate 200iC operate the real automated workplace (b) [source: own study]

4. CONCLUSION

Automation of manufacturing technologies with industrial robots is one of ways how to increase the production quality and productivity. Industrial robots as flexible automatically working machines are able to adapt themselves on the changes in the product production very fast by the flexible reprogramming way. Utilisation of computer aid, namely of simulation offline programs for robots programming doesn't cause a downtime in production by programming and shorts starting of new product production. The article presents the possibilities of the Roboguide HandlingPRO utilisation for the solution of partial tasks of automation of manufacturing technologies with robots on the real examples. Our practical experience has shown that the use of appropriate simulation tools of Roboguide HandlingPRO can help to verify the proposals of robot workcells for automation of welding, deburring and assembly technologies.

ACKNOWLEDGMENTS

This article was made with support of project KEGA 037ŽU-4/2014 – The Development of Complex Interactive Educational Portal for Support the Teaching of CNC Production Machines Programming.

REFERENCES

- [1] CUBONOVA N., KURIC I.: *Data structures implementation of the protocol STEP-NC at CNC machines programming*. Communications: scientific letters of the University of Žilina. Vol. 16, No. 3A, 2014, pp. 176-183, ISSN 1335-4205.
- [2] ČUBOŇOVÁ N.: *Počítačová podpora programovania CNC strojov*. Computer Aid of CNC Machines programming, Žilinská Univerzita, Vyd. 1, Žilina 2012, p. 114, ISBN 978-80-554-0514-8.
- [3] Delfoi Robotics: http://www.delfoi.com/web/solutions/robotiikka/en_GB/offline/
- [4] KUBO M.: *Návrh riešenia automatizácie zváracieho pracoviska s využitím zváracieho robota. (Design of the solution for automation of welding workplace with usage of welding robot)*. Diploma thesis. University of Žilina, FME, Department of automation and production systems. Supervisor: Assoc. Prof. Eng. Darina Kumičáková, PhD., 2014, p. 69.
- [5] KUMIČÁKOVÁ D.: *Využitie prostriedkov počítačovej simulácie pri navrhovaní automatizovaných montážnych pracovísk. (Utilisation of computer simulation tools in automated assembly workplaces designing)*. ATP journal plus [elektronický zdroj]: recenzované periodikum vedeckých a inžinierskych publikácií. CD-ROM. Č. 1, 2013, pp. 17–19, ISSN 1336-5010.
- [6] KUMIČÁKOVÁ D., ZAHORANSKÝ R.: *Laboratory workplace of automated assembly – contribution to the education quality improvement*. Materials science and technology, Mim. č., 2010, [elektronický zdroj]: internetový časopis. pp. 34–40, ISSN 1335-9053.
- [7] KURIC I. et al.: *Počítačom podporované systémy v strojárstve. (Computer Aided Systems in Mechanical Engineering)*, Žilinská univerzita, Vyd. 1, Žilina 2002, p. 351, ISBN 80-7100-948-2.
- [8] ROBOGUIDE – FANUC Simulation Software. Available on: <http://robot.fanuc-america.com/products/vision-software/roboguide-simulation-software.aspx>
- [9] URÍK M.: *Konceptný návrh tréningového laboratórneho pracoviska pre robotizované obrábanie. (Conceptual design of training laboratory workplace for robotized machining.)*. Diploma thesis. University of Žilina, FME, Department of automation and production systems. Supervisor: Assoc. Prof. Eng. Darina Kumičáková, PhD., 2014, p. 74.
- [10] YAHUI G., IANZHONG D., DONGWEI L.: *Off-Line Programming Techniques for Multirobot Cooperation System*. International Journal of Advanced Robotic Systems., Vol. 10, 2013, 282:2013. p.17. Available on: <http://cdn.intechopen.com/pdfs-wm/45451.pdf>

Patryk SOKOŁOWSKI*, Jacek CZARNIGOWSKI**,
Paweł MAGRYTA***

CFD SIMULATION STUDY OF AIR FLOW AROUND THE AIRFOIL USING THE MAGNUS EFFECT

Abstract

The article presents a simulation study of air flow around the airfoil equipped with an additional rotary element, using the Magnus effect to change the air flow around it. For this purpose, a two-dimensional computational model for the airfoil without and with the moving element was made. The model was based on the airfoil NACA 8-H-12. The article describes a method of creating a model in two versions: without moving computational grid and with moving computational grid. Studies were carried out for a typical flow rate corresponding to the flow velocity around the gyroplane in a nominal flight speed conditions. Article shows a comparison of flows in both cases, and assesses the impact of the geometry that use the Magnus effect on the work of airfoil.

1. INTRODUCTION

In the year 1853 H.G. Magnus observed and discovered the fluid dynamics phenomenon of existing the force that is perpendicular to the direction of body motion and acting on the rotating roller or other rotary body moving relative to the fluid (Fig. 1). Magnus effect can be often observed in sports. By using this aerodynamic effect, soccer player can spin the ball properly to shoot it into the net. The golf ball after appropriate shot by the golfer can increase or decrease the

* Lublin University of Technology, Faculty of Mechanical Engineering, Department of Thermodynamics, Fluid Mechanics and Aviation Propulsion Systems, Lublin 20-618, Nadbystrzycka 36, p.sokolowski@pollub.pl

** Lublin University of Technology, Faculty of Mechanical Engineering, Department of Thermodynamics, Fluid Mechanics and Aviation Propulsion Systems, Lublin 20-618, Nadbystrzycka 36, j.czarnigowski@pollub.pl

*** Lublin University of Technology, Faculty of Mechanical Engineering, Department of Thermodynamics, Fluid Mechanics and Aviation Propulsion Systems, Lublin 20-618, Nadbystrzycka 36, p.magryta@pollub.pl

distance at which it can fly. These situations are caused by the fact, that on both sides of the ball, differential pressure is created and where the ball is spinning, the return force is directed opposite to the force of gravity, what makes the ball flight extension.

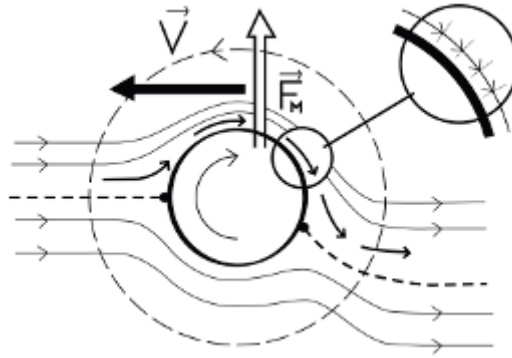


Fig. 1. The principle of Magnus effect on the example of ball [1]

Ballistics is also the area where the Magnus effect has a very big impact on the track and the trajectory of the bullet flight. In the case of using threaded barrels, the shooter, especially snipers and marksmen, but also a modern fire control systems in the third-generation tanks, have to take an amendment to the wind direction relative to the axis of the barrel.

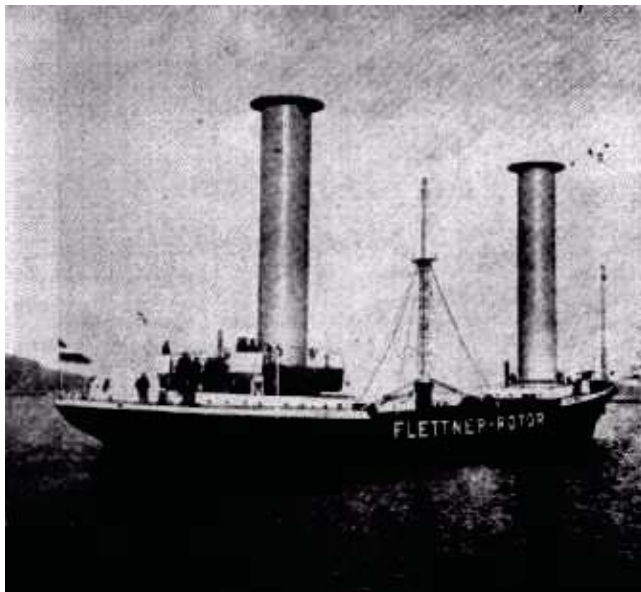


Fig. 2. Flettner rotor driven ship [2]

In 1920, German engineer A. Flettner developed a new type of ship engine, that used the principle of the Magnus effect. The drive was constructed from one, two or three rotating cylinders. As a result of cylinder rotations, the Magnus force was produced, what allows the movement of the ship (Fig. 2).

Magnus effect is also used in the case of renewable sources of energy – wind turbines. Magnus Effect turbine blades have classic cross-section shape of airfoil that consist of cylindrical rotor, which as a result of produce lift allows the rotor to rotate the whole turbine. Magnus effect was is also observed in subsonic and supersonic flight [3].

The rotating elements makes sufficiently large flow changes around its geometry that can be used to change the air flow in the whole area. What was proven in scientific papers [4]. The question therefore arises: Whether this phenomenon can be used to change the air flow around the airfoil or improving its aerodynamics? This article presents a preliminary analysis of the concept of airfoil design for gyroplane use, with the additional Magnus Effect.

2. RESEARCH OBJECT AND RESEARCH PLAN

Airfoil [1] used for the simulation is the NACA 8-H-12. This airfoil is used in the construction of the main rotor of gyroplane in the Department of Thermodynamics, Fluid Mechanics and Aviation Propulsion Systems at the Faculty of Mechanical Engineering in Lublin University of Technology. Airfoil structure was modified by adding the cross rotor and placing it in distance of a 40% of chord length of the airfoil. Chord length is 100 mm. During the rotation of the rotor its arms protrude beyond the outline of the airfoil by of the length of 0.5 mm (Fig. 3).

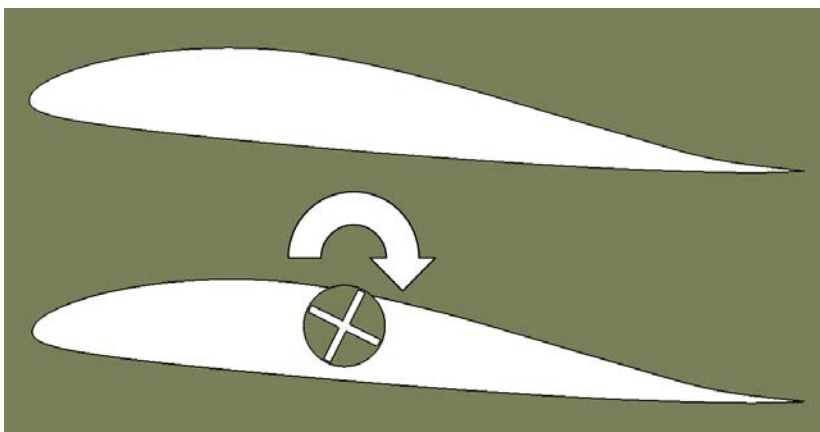


Fig. 3. Geometry of the airfoil without and with the moving rotor design in the Catia V5 [source: own study]

Simulation studies were carried out for the two-dimensional model. Airfoil was set at the same angle of attack of 5° . The flow velocity was set at 23.5 m/s, the direction of the rotor rotation was consistent with the clockwise direction (fig. 3), the speed of the rotor rotation was 1000 rpm. The flow velocity of the model corresponds to 500 rpm in the middle distance of the main rotor blade of gyroplane (diameter of 1.8 m). So the speed of the rotor rotation correspond to the twice speed of the entire gyroplane main rotor. Geometry prepared in this way the was exported to ANSYS Mesher environment for preparation of computational grid.

3. MODEL PREPARATION

3.1. Geometric model

The geometries of the two designs were developed in CATIA v5 [5]. Models based on the data entered in the form of airfoil individual points coordinates. Developed geometry was exported to ANSYS Geometry module. In both cases, geometry was divided along the airfoil chord for the two bodies for better generation of computational grid (fig. 4). In the case of the second airfoil, rotor has been imported into the Geometry module as a separate body, due to the rotation process.

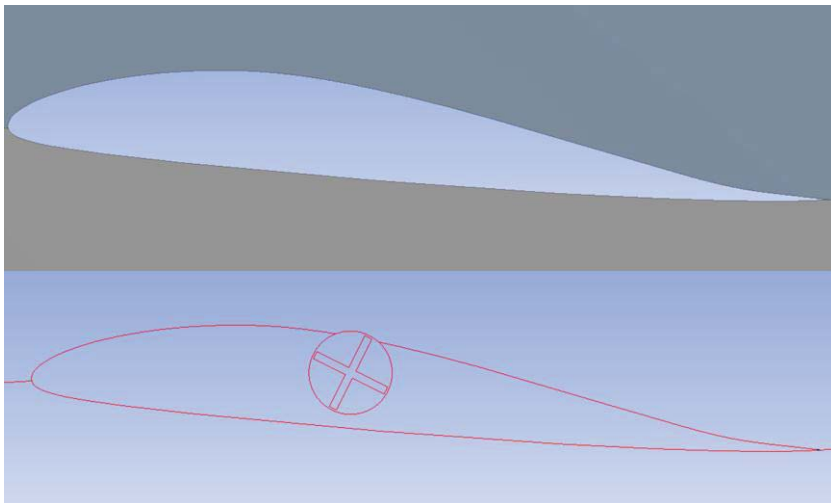


Fig. 4. Geometry imported into ANSYS Geometry module [source: own study]

3.2. Computational model preparation

The first step in the preparation of a discrete model to perform calculations, was to create a computational grid. Model prepared in Geometry module was transferred to the Mesher module [6], which is used to generate a computational grid. Below there is a description how computational grid was generated, on the example of the airfoil with the moving rotor.

Tetrahedral computational grid was selected due to its ease of preparation and control of the boundary layer. Therefore, for the entire geometry a Method function has been attributed, where the type of tetrahedral grid is selected. Then, the edges of the airfoil contour has been concentrated using Sizing option with the setting of the size of elements of 0.25 mm (Fig. 5).

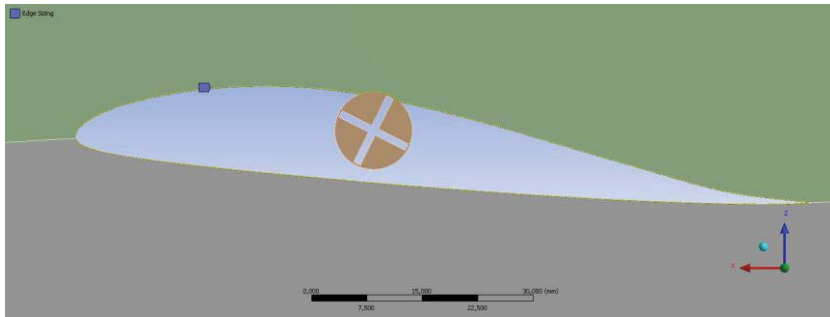


Fig. 5. Edge sizing on the airfoil contour [source: own study]

Tetrahedral grid type, as opposed to a regular grid, requires the creation of additional elements that will be responsible for the boundary layer on the tested airfoil. In the Mesher module, an Inflation option is responsible for creating a boundary layer in the computational grid.

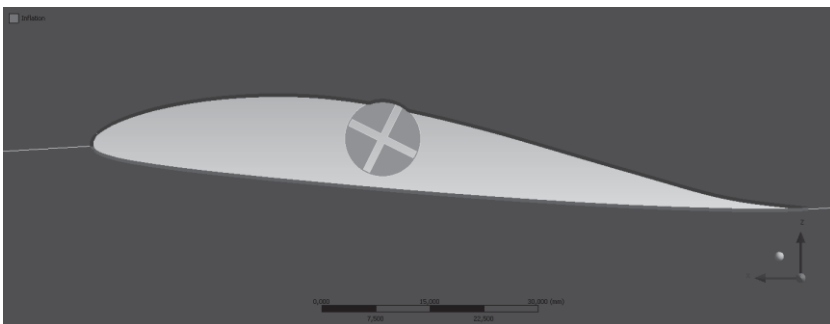


Fig. 6. Inflation option shown in the airfoil area [source: own study]

As can be seen in Figure 6, to fully activate the Inflation option, it is necessary to choose the space where the geometry of the boundary layer has to be created (blue area) and the limit of the boundary layer - in this case, the airfoil contour (red area).

Similar boundary conditions for computational grid have been given to the geometry of the rotor, but as Sizing option, a Body Sizing was chosen, by limiting the size of the elements for the entire body to 0.25 mm. Boundary layer conditions have also been given to the shoulders of the rotor and the edges of the circle (fig. 7).

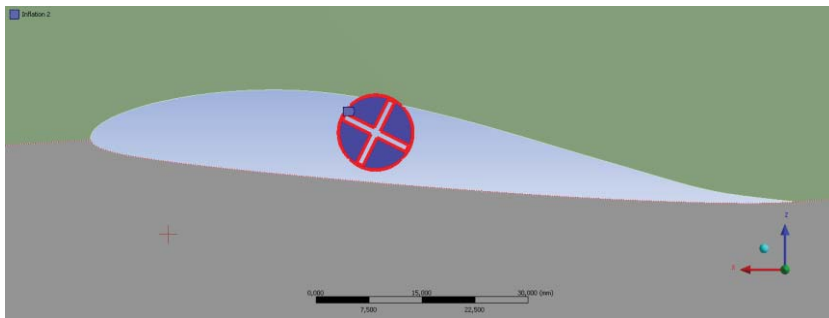


Fig. 7. Inflation option shown in the rotor area [source: own study]

When defining the computational grid, Named Selection option should also be given on items that will be used to give the boundary conditions. Such relations were given to the edges indicating points of inlet and outlet of the fluid flow (Fig. 8).

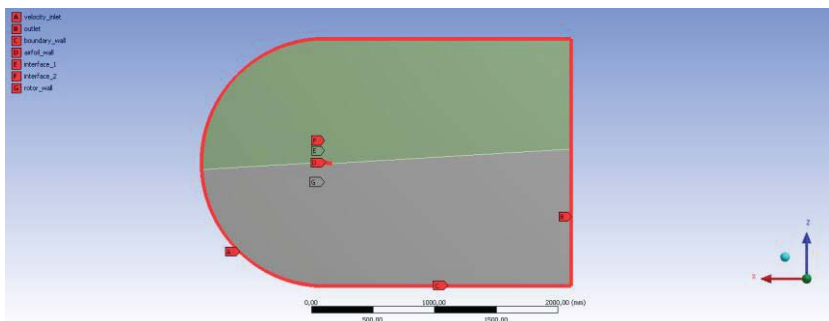


Fig. 8. Named Selection in computational grid [source: own study]

The edges of the contact between the rotor and the airfoil is also indicated by the Named Selection option by using the names: interface_1 and interface_2. In this way, after defining Interface option, fluid flow between the two bodies has been enabled (Fig. 9).

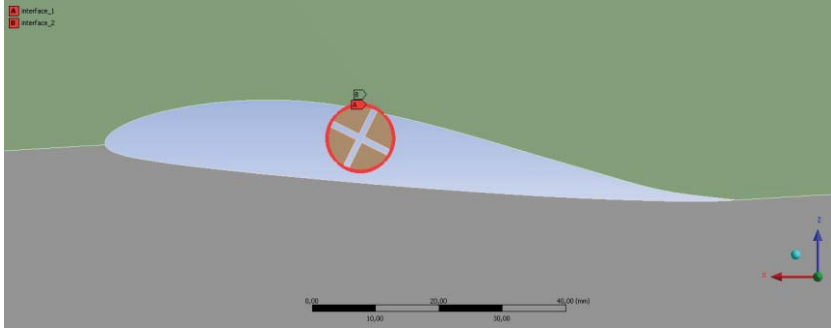


Fig. 9. Interface given on the edge of two bodies [source: own study]

For the preparation of computational grid for the airfoil geometry without the moving element, Interface definition of the relation type and steps that define the grid options should be omitted. Figure 10 shows generated computational grid for the case with a moving geometry.

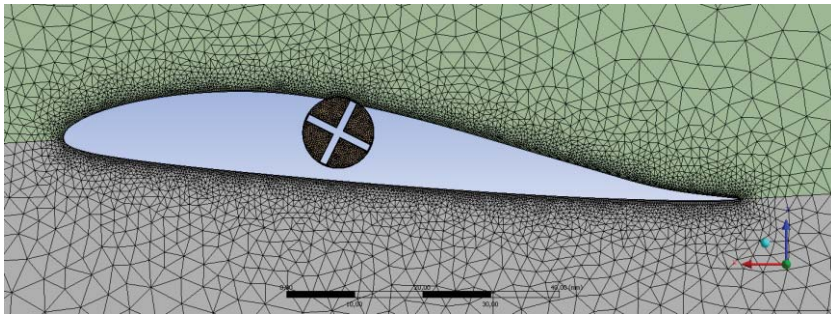


Fig. 10. Generated computational grid around the airfoil of the moving geometry [source: own study]

Prepared computational grids have been loaded into the ANSYS Fluent software, where the boundary conditions were given. Simulation for the geometry with the moving element has been defined as Time: Transient, but for the 2D simulation of airfoil without moving geometry, the simulation was set as Time: Steady. For the turbulence model k-omega SST with enabled energy equation was selected. Liquid material was set as incompressible ideal gas (air). Turbulence model and the type of fluid in the two simulations have been the same. There was also evaluation of the turbulence model [7].

Due to the movement of the rotor, the right conditions to allow this movement should have been given. By creating a separate body of a circle containing the rotor (Fig. 11), it was possible to set Mesh Motion option in Cell Zone Conditions for the body of the rotor (Fig. 12).

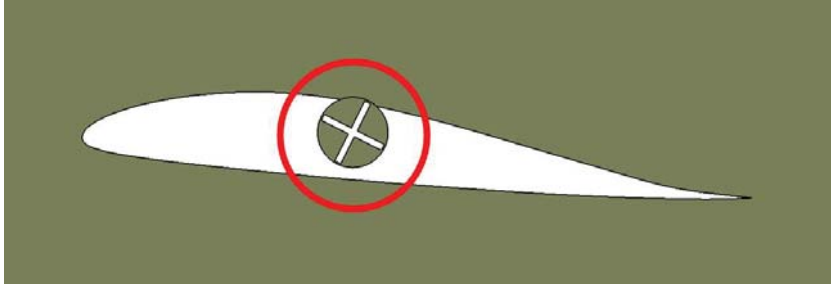


Fig. 11. Specified body containing a rotating part [source: own study]

Zone Name
rotor

Material Name air Edit...

☐ Frame Motion ☐ Laminar Zone ☐ Source Terms
☒ Mesh Motion ☐ Fixed Values
☐ Porous Zone

Reference Frame Mesh Motion Porous Zone Embedded LES Reaction Source Terms Fixed Values Multiphase

Rotation-Axis Origin
X (m) 0 constant
Y (m) 0 constant

Rotational Velocity
Speed (rpm) -1000 constant

Translational Velocity
X (m/s) 0 constant
Y (m/s) 0 constant

Relative Specification
Relative To Cell Zone absolute

Copy To Frame Motion

UDF
Zone Motion Function none

Fig. 12. Settings dialog box for Mesh Motion option [source: own study]

The next step was to add interface between the two bodies to allow free flow of liquid during the simulation. Because of the transient nature of the simulation and simulation of the moving element, it was required to determine the time step and the amount of time steps. Time Step was set to 0.025 s, and the number of time steps was set to 200. These settings correspond to 5 seconds of flow around the airfoil. At each time step the number of iterations was defined to 50.

4. SIMULATION STUDIES

The model allowed for the execution of simulation studies in the target range. We analyzed the distribution of pressure and velocity around the airfoil for the two cases examined and the fluid flow around airfoils.

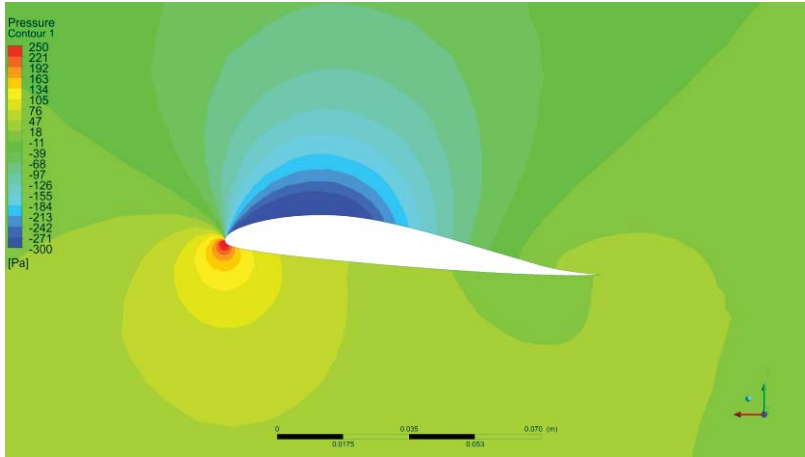


Fig. 13. Pressure distribution around airfoil without moving element [source: own study]

Figure 13 shows the distribution of the pressure around the airfoil. In the figure, it can be seen that near airfoil nose there is a high pressure (red area). This area is associated with the direction of air flow. On top of the airfoil a low pressure region can be seen (blue area), the maximum under pressure in this area is approximately 350 Pa. Due to the design of airfoil and the differences in pressure under the desired conditions, lift force directed in the direction of the Y axis is generated.

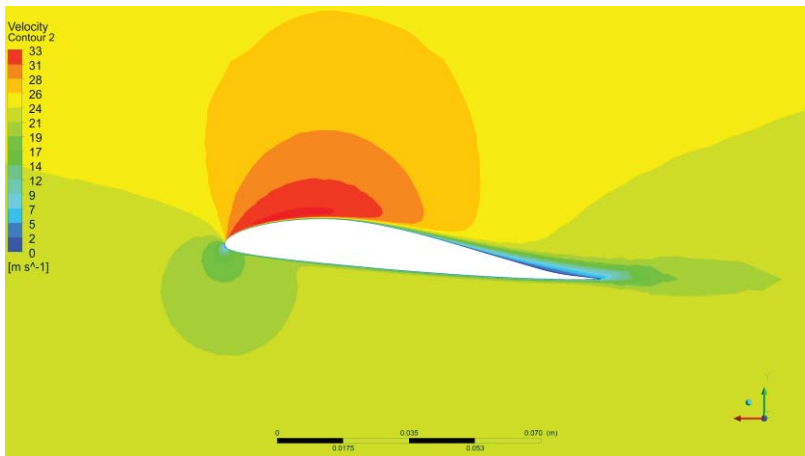


Fig. 14. Velocity distribution around airfoil without moving element [source: own study]

On the distribution of velocity around airfoil, it can be seen that the area where there was a high pressure, there is also a low velocity area of the air (Fig. 14).

While on the upper edge of the airfoil a area of high fluid velocity occurs, velocity reaches the speed of 33.5 m/s.

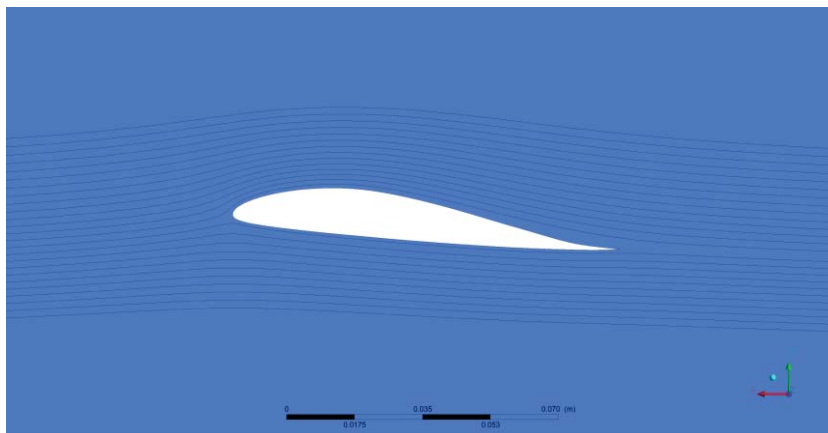


Fig. 15. Streamline around airfoil without moving element [source: own study]

Figure 15 shows the streamline around the airfoil. For visualized streamline, vortex around the airfoil cannot be seen, but comparing both, the velocity distribution and streamlines, it can be noted that, in the vicinity of 60-70% of the chord length, the fluid separates from the top of the airfoil. The tested airfoil is characterized by a drag force 0.94 N acting against the movement of the fluid in the X-axis and the lift force 21.40 N acting in the direction of the Y axis.

Figure 16 shows the results of the simulation for the airfoil with moving rotor for the five time steps – for 1, 2, 3, 4, and 5 seconds of the simulation time. The left column shows the pressure distribution for each of the selected time steps, and the right column shows the velocity distribution for each of the time steps.

Pressure distribution around presented airfoil is similar to the airfoil without the moving element, it is characterized by two areas of high and low pressure. An area of low pressure, depending on the position of the rotor, decreases or increases, what allows to notice that both the lift and the drag forces may be characterized by a period decreases and increases. This situation is illustrated in figure 17 representing the changes of the drag force acting on the airfoil depending on the simulation time step.

Figure 18 represents the changes of the lift force acting on the airfoil depending on the simulation time step. On figure 16 – the distribution of velocity and figure 19 – streamlines around the airfoil, it can be noticed that at 2, 3 and 5 second there is a clear area of detachment of the liquid from the airfoil starting at about 45% chord length of the airfoil. This suggests area with very low fluid flow rate in this part. The streamline visualization shows also an increased distance between the streamline and airfoil in this area.

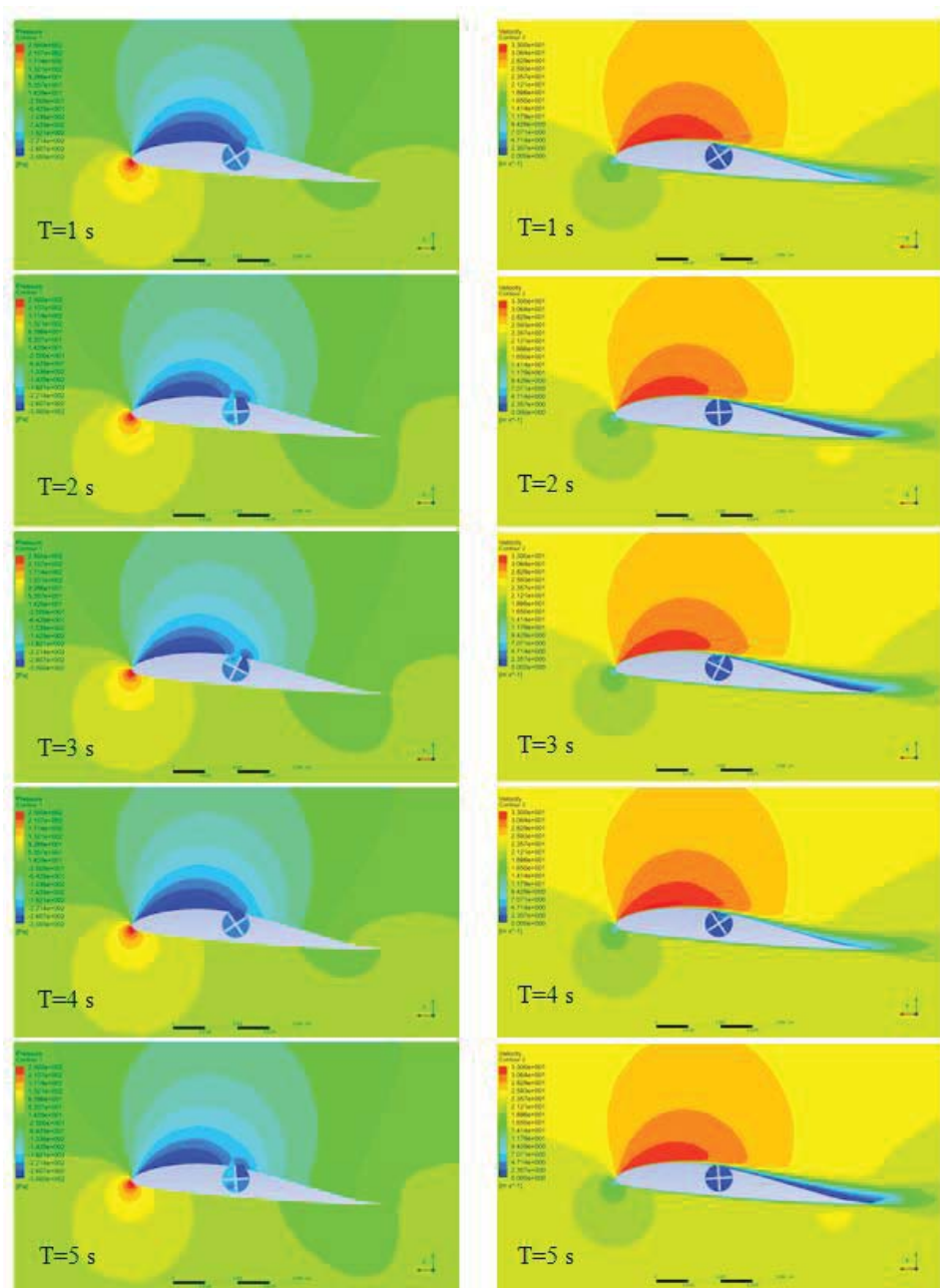


Fig. 16. Pressure and velocity distribution around the airfoil with moving rotor [source: own study]

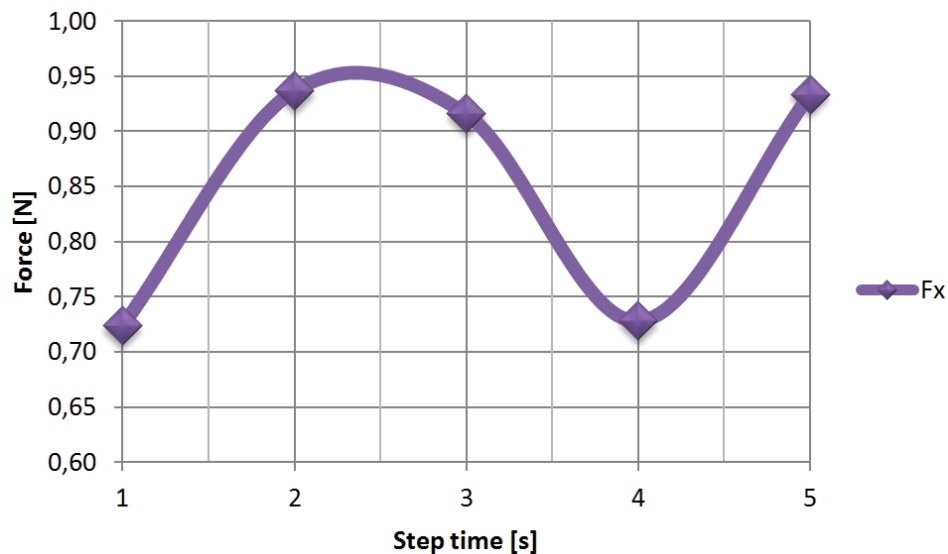


Fig. 17. Changes of the drag force acting on the airfoil depending on the simulation time step [source: own study]

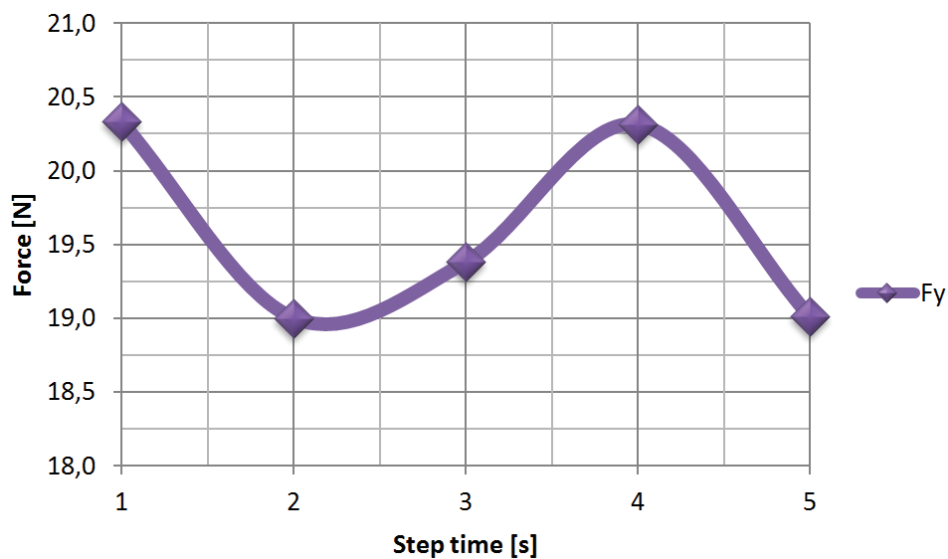


Fig. 18. Changes of the lift force acting on the airfoil depending on the simulation time step [source: own study]

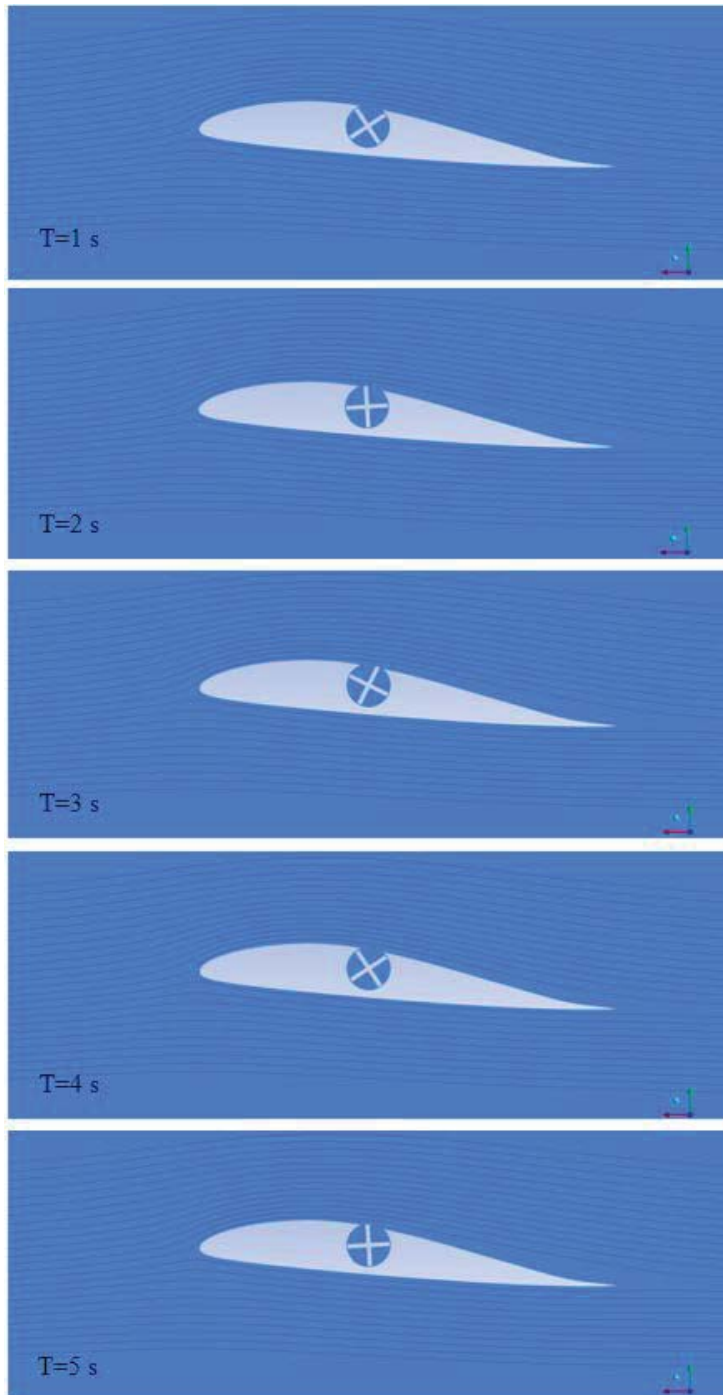


Fig. 19. Streamline around airfoil with moving element [source: own study]

Tab. 1. Summary of the forces in next time steps

	Step Time [s]					Average
	1	2	3	4	5	
Drag Force [N]	0,72	0,94	0,92	0,73	0,93	0,85
Lift Force [N]	20,33	19,00	19,38	20,31	19,01	19,60

Comparing the velocity distribution and streamlines data in table 1, it can be seen that in these steps when the rotor arms protrude beyond the outline of the airfoil, drag force acting on the airfoil increases compared with the 1st and 4th step by 30%, while the lift force decreases by 7%, which adversely affects the aerodynamic characteristics of the airfoil.

5. SUMMARY

Analyzing the results it can be concluded that the developed models allowed us to determine the difference in air flow around a airfoil with and without the additional moving element. The use of such a device meant that due to the small number of rotor arms, the forces acting on the airfoil change periodically which adversely affect the operation of airfoil for a main rotor of gyroplane. The average drag force acting on the airfoil is lower than the drag force acting on the airfoil without rotor by 10%, which is a positive result. Average lift force is less than the force acting on the airfoil without moving element by 8.5%. It follows that the airfoil aerodynamic balance has improved and its perfection increased by 1.3%.

Despite the reduction in the drag force acting on the airfoil with the moving rotor and improve in the aerodynamic excellence, proposed design, in its current form is not possible. It should be changed, by changing the geometry of the rotor and the elimination of sinusoidal forces affecting on airfoil and also optimize the position of the rotor in airfoil contour.

REFERENCES

- [1] BUKOWSKI J.: *Mechanika Płynów*. Warszawa, PWN, 1970.
- [2] ŁOPACIUK P.: *Wizualizacja efektu Magnusa z wykorzystaniem biblioteki OpenGL*. Diploma Thesis, Chełm, 2007.
- [3] JAMES DESPIRITO WEAPONS AND MATERIALS RESEARCH DIRECTORATE, ARL: *CFD Prediction of Magnus Effect in Subsonic to Supersonic Flight*. Army Research Laboratory, Aberdeen Proving Ground, MD 21005-5066.

- [4] SEDAGHAT A., SAMANI I., AHMADI-BALOUTAKI M., ASSAD M., GAITH M.: *Computational study on novel circulating aerofoils for use in Magnus wind turbine blades*. Energy 91, 2015, pp. 393-403.
- [5] *Catia V5 – Program Help*.
- [6] *ANSYS Workbench – Program Help*.
- [7] ELENI D., ATHANASIOS T., DIONISSIOS M.: *Evaluation of the turbulence models for the simulation of the flow over a National Advisory Committee for Aeronautics (NACA) 0012 airfoil*. Journal of Mechanical Engineering Research, Vol. 4, No. 3, 2012, pp. 100-111.

*Irina KUSHNIRETSKA**

FORMING OF THE SEMISTRUCTURED DATA DYNAMIC INTEGRATION MASH-UP SYSTEM CONTENT

Abstract

This paper describes the method of forming a united dynamic data set that has the general structure and only content. The procedure of forming the triplets with the structure "subject-predicate-object" with the received input information resources descriptions has been proposed. The formula of calculating the similarity factor of user query with information resource semantic metadata has been presented. The structure of the semistructured data dynamic integration system that used "Mash-Up" technology has been designed.

1. INTRODUCTION

Participation of end-users is an essential driving force of web technologies developing. Although the application of Web 3.0 allows the spread of Internet using to more and more areas, is still remain the controversial issue: how not specialized ordinary users can interact with them and be more than just a receiver of data. The current state of Web 2.0 already provides opportunities for end users to evolve from simple consumers of information in information developers. This allows the new data integration systems working using Mash-Up technology. Currently, such systems are actively developing and allow users to collect, transmit and use Web resources. The purpose of these systems is to create new and useful applications from available web resources.

However, existing methodologies and tools for software systems building are focused on the well-structured problem with sufficiently formalized subject areas and permanent local knowledge sources. So many problems associated with providing semantic of information during the semistructured data Mash-Up-system dynamic integration remain unresolved.

* Lviv Polytechnic National University, Ukraine, 79013, Lviv, Bandera str., 28a, presty@i.ua

2. THE PROBLEM FORMULATION

The importance of various semistructured data integrating, today, hard to overestimate. The ability to quickly and finding the necessary quality information to make informed decisions adequate and necessary information required to ensure the process of storage, storage, analysis and interpretation of all required data. Using web-search tools you can find a huge number of diverse

Information specific topics, but not always achieved with the necessary coherence and consistency of the data. Integrity, consistency, data consistency can be effectively achieved only if the use of special methods of information processing single centralized system regardless of the architecture and implementation of information infrastructure organization. Building such infrastructure is lengthy and time consuming process, the complexity of which depends not only on the volume of accumulated historical data, as the number and diversity of sources and different applications. Therefore, crucial in this process certainly plays the correct choice of methods and means of data integration.

Hence, the aim of this work is to research of the application of new approaches for solving the problem of increase the quality of semistructured information received from various web-systems storage and presentation by forming a united dynamic data set that has the general structure and unique content.

3. THE BASIC MATERIAL PRESENTATION

3.1. Dynamic data integration systems based Mash-Up technology

Recently Mash-Up technology became a trend which allows non-professional users to create Web applications, combining functionality with more than one with the important task during Mashup system creating is data getting Web-services to solution the situation and specific tasks.

Mashup-systems often focus on one or more of the following objectives:

- Extract data from existing data sources such as Web-pages, Web-services and channels.
- Data combining from different sources into a single data set. The data from each source must have at least one general area, for example, name of the restaurant, so that one set of records could be matched with corresponding entries in the other sets.
- Data visualization in a way that allows the user to understand the aggregated data. Many collages include geographical data, such as addresses (visualized using Google Maps).

Unfortunately, the general classification of Mashup applications is absent. Because one of the most important tasks when Mashup system creating is data getting, it is wise to make a classification of the data type with which Mashup system operates. There are four main categories: maps; media content, video and photos; news; search and purchase.

The main objectives for the Mashup tools searching [4]:

1. Achieving more efficient Mashup construction;
2. Achieving of integration of Mashup systems developing as part of the software development process;
3. Conduct research and design Mashup existing tools.

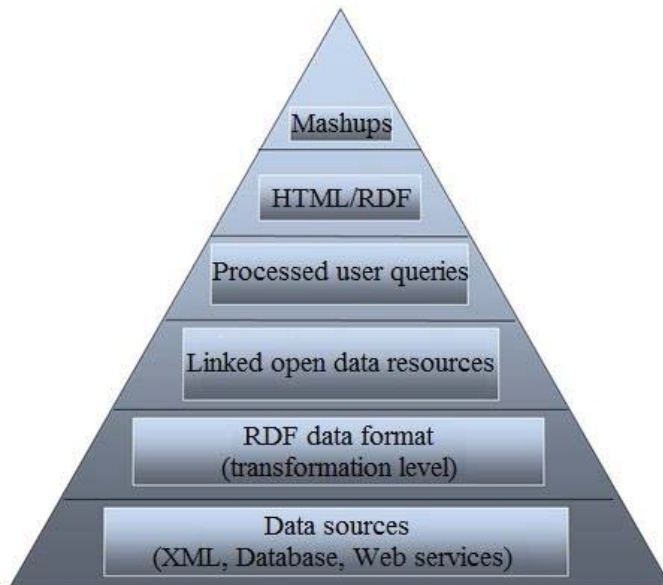


Fig. 1. "Pyramid values" of the semistructured data dynamic integration system based on "Mash-Up" technology

Today we know a number Mashup systems that help the user to create the data mashups. We chosen the following systems for three main reasons:

1. These systems are most popular when this analysis was performed (based on discussion forums, blogs, etc.).
2. Information about some other systems could not be complete because there is unavailability of systems data at certain stages of research, not allowing us to experiment and report results according to our analysis.
3. The reason for this is motivated by the fact that our aim is not to analyze all systems, but provide an indication of the current state of these systems and understand their overall approach to data integration.

IFTTT [5] ("if this, then that") – Mashup service that allows users to connect to various web applications using simple conditional statements, known as "recipes" and create simple automated sequence that runs when performing an action. IFTTT was developed by the American programmer Linden Tabetson and launched in 2010. IFTTT allows users to create and share "recipes" that correspond to the judgment: "if this, then that" "it" part of the recipe, called trigger. Pretty simple to use and consists of only three tabs:

- Tasks – a list of your active tasks.
- Recipes – a list of the most popular tasks that you can use as their own, that is something like pieces tasks.
- Channels – a list of supported services at the time of the study 54 (e.g., Facebook, Twitter, LinkedIn, RSS, Google RSS Reader, Evernote, Gmail, Google Calendar, WordPress, etc.).

Example working with IFTTT: When we update status on Facebook, want it to be immediately twitt on Twitter. In addition to its Web application, released a mobile version for iPhone service in July 2013. Android-version of the application was released on April 24, 2014.

One with the worthy analog IFTTT can be called a service called Zapier [6], which was created in 2012 by the American development team: Brian HelmiHom, Wade Foster and Michael Knop. It fully inherited the trigger-action circuit and allows you to build relationships between different web applications. Unlike IFTTT Zapier supports 147 channels, 2 times more than ifttt, and also its interface for many people seem more intuitive and simple. But unlike the completely free IFTTT, Zapier has 4 packages - one free and three paid with a limited number of services, tasks and trigger's actions for each packet.

Yahoo pipes [7] – web-tool provided by Yahoo. Users can create collages by aggregating and processing of web-channel, web-pages and other services. Pipe consisting of one or more modules, each of which performs a certain task, for example, such as: receiving data from web-sources, filtering, sorting or merging channels. The resulting system data can be accessed by the client with a unique URL as RSS or JSON, or through visualization on Yahoo map.

Google Alerts [8] – service from the search giant. It is based on the idea of query results monitoring according to the time changing. In fact, you can set up "alerts" to new results for. The system is able to filter out "all rubbish" and select the most relevant data.

So, in order to get the Mashup system work quality results should provide the necessary for this integration, coherence and consistency of the data. So critical for this problem solving is certainly in the right choice of methods and means of data receiving, storing and presenting.

3.2. Materials and research methods of the combined dynamic data set content forming that has the general structure and united content

For received input information storing in the system in a structured way useful will be using such semantic-oriented technologies such as ontology and description logic [9].

According to [10], ontology based on description logic is a signs system:

$$O_{DL} = \langle C, CD, R, A, I, V, R_I, A_I, L, P_C, P_R, P_A, P_{IC}, P_{LC}, P_{LR}, P_{LA}, P_{LI} \rangle \quad (1)$$

where: $C = \{c_1, \dots, c_n\}$ – concepts finite set in ontology,
 $CD = \{cd_1, \dots, cd_k\}$ – standard data types set, including two types {string, integer},
 $R = \{r_1, \dots, r_m\}$ – binary relations final set between concepts,
 $A = \{a_1, \dots, a_w\}$ – finite set of attributes (binary relations between concepts and standard data types),
 $I = \{i_1, \dots, i_z\}$ – finite set of instances in the ontology,
 $V = \{v_1, \dots, v_q\}$ – standard type specific values finite set,
 $R_I = \{ri_1, \dots, ri_m\}$ – concretized relations finite set (binary relations between instances) $ri_I(i_x, i_y)$,
 $A_I = \{ai_1, \dots, ai_w\}$ – concretized attributes finite set (binary relations between instances and i specific values $ai_I(i_x, v_y)$),
 $L = \{l_1, \dots, l_f\}$ – lexical label final set (ontology dictionary),
 $P_C \subseteq C \times C, P_C \in R$ – anti symmetric, transitive, anti reflexive binary relation, which is the relation of the partial order on the concepts set C,
 $P_A \subseteq A \times A$ – anti symmetric, transitive, anti reflexive binary relation, which is the relation of the partial order on the attributes set A,
 $A = \{a_1, \dots, a_w\}$ – finite set of attributes (binary relations between concepts and standard data types),
 $P_R \subseteq R \times R$ – anti symmetric, transitive, anti reflexive binary relation, which is the relation of the partial order on the relations set R,
 $P_{IC} \subseteq I \times C$ – incidence binary relation between sets I and C,
 $P_{LC} \subseteq L \times C$ – incidence binary relation between sets L and C,
 $P_{LR} \subseteq L \times R$ – incidence binary relation between sets L and R,
 $P_{LA} \subseteq L \times A$ – incidence binary relation between sets L and A,
 $P_{LI} \subseteq L \times I$ – incidence binary relation between sets L and I.

The ontology definition in formula (1) is given with regard to the description logic properties.

When the user's query analyzing and answer to it forming in the form of information resources integrated set there arises the problem of comparison on the similarity of the query and integrated data. To solve this problem is proposed to use the similarity strings metric. Currently there are a number of commonly used universally metrics of determining the similarities of two text strings [11]: Lowenstein distance or edit distance, Zhakkar-Winkler coefficient, Tanimoto coefficient and Severensen-Dicey coefficient.

Let X and Y - two strings of length m and n. For Lowenstein distance getting are calculated the distances matrix D, in which each element D [i, j] contains the distance between the first character I of string X and the first character j in the string Y. Lowenstein distance is determined by the following formula:

$$D(i, j) = \left\{ \begin{array}{l} \max(i, j), \text{ if } \min(i, j) = 0 \\ \min \left\{ \begin{array}{l} D(i, j-1) + 1 \\ D(i-1, j) + 1 \\ D(i-1, j-1) + m(X[i], Y[j]) \end{array} \right\} \end{array} \right\} \quad (2)$$

where: $m(a, b) = 0$ if $a = b$,
 $m(a, b) = 1$ if $a \neq b$.

Tanimoto coefficient or Zhakkar-Winkler also often are used to determine the similarity of one string to another. In this case, the string is seen as a set of characters, and the similarity metric determines the amount of the same characters as follows:

$$p = \frac{c}{a + b - c} \quad (3)$$

where: p – strings similarity factor: $0 \leq p \leq 1$,
c – number of joint characters in strings,
a and b – number of characters in strings A and B respectively.

By the same principle operates Severensen-Dicey coefficient, which is a binary measure of strings similarity and has in general the following entry:

$$p = \frac{2 \times n(X \cap Y)}{n(X) + n(Y)} \quad (4)$$

where: p – strings similarity factor: $0 \leq p \leq 1$,
 X and Y – compared strings,
 $n(a)$ – function that determines the number of characters in the string a .

3.3. The method of the combined dynamic data set, which has the general structure and united content forming

In [12] based on the Mashup systems activity analysis are the following activity states:

1. Registration. When there is the successful registration – move to the second step, the unsuccessful registration – returning back to the registration beginning.
2. Authorization, if all goes well – move on, if not – back to the authorization beginning.
3. Task formulation. If the task is formed according to the rules of the system, we move on, if not – return to the beginning of the third step.
4. Mashup system sources forming.
5. In selected sources of the required information finding. If the results are satisfactory – we go further, if not – return back to the searching.
6. Relevant information extracting and to the next state transition.
7. Received information storing in the service form.
8. The finished Mashup system result visual presentation.

The most important states in Mashup system work, according to [12], are: required information finding (fifth state), found information extracting (sixth condition) and storage as a service (seventh state). To improve of the result receiving for seventh state work in [12] the procedure of determining the structure and content of incoming information resources has been proposed. This procedure can be interpreted as a method of determining the structure and content of the received input information. The usage of this method makes it possible to increase the quality indicators of the fifth state result. And, as every subsequent state depends on the previous one – it increases of the two next important activity states productivity.

User request analysis and answer forming to it – some of the tasks that need to be solved at the seventh state of Mashup system work. To improve of seventh state (storing as a service) work result quality indicators we propose a method of forming a united dynamic data set that has the general structure and unique content.

The method of forming the united dynamic data set that has the general structure and unique content consists of the following steps-tasks:

Step 1: Analysis of the received metadata information resources.

If semantic metadata is generated based on a textual description of the object and no clear selection of concepts and instances, then the data will process according to the second step tasks. And if all concept and instances meta description are selected, then the data will process according to the third step tasks.

Step 2: Searching of the concepts and instances in the information resources textual descriptions.

The textual description is analyzed for the presenting of concepts and instances that can serve as subjects in the semantic metadata elements. To solve the problem at this step was used the instances and concepts in the text search function. The objective of this function is to search of lexical labels of concepts and instances with ontology in the object textual description to form a set of possible subjects in the semantic metadata elements. Of course, the result of this function can't completely replace human work and after using search function of concepts and instances you have to use of additional technologies for:

- to remove of all objects that do not reflect the essence of the object description;
- to eliminate multiple meanings if the set contains items with the same lexical labels;
- to complete the set of concepts and instances that were not found as a result of the function.

Then the elements of set can be used to form triplets according to certain rules of predicates and objects selection that we describe a little later.

Step 3: Forming of triplets with the structure "subject-predicate-object" with received semantic information resource metadata.

We need that the received our semantic metadata items be in form of the triplets with the structure "subject-predicate-object" or separate concepts or instances of ontology, which will be called the "subject". In order to structure your content semantically is necessarily need to be given "subject".

If the subject is given for the essence of the subject description presentation, then on the selection of predicate and object are imposed additional restrictions that arising from the rules of formation of the description logic statements.

The set of possible predicates in triplet is limited by chosen triplet subject. That is, as a predicate can be selected those relations or attributes, that in ontology are defined for the subject - concept or instance.

After predicate selecting is necessarily indicate the triplets object. The set of possible objects depends on the selected predicate. The rules of forming the set M_o of possible objects in triplets based on the definition of ontology in formula (1) are as follows:

1. If the predicate value – is the relation r_x , then

$$M_o = \{o_i \in C \cup I \mid r_x(c_x, o_i) \vee (r_x(c_x, c_y) \wedge P_{IC}(o_i, c_y))\} \quad (5)$$

2. If the predicate value – is the attribute a_x , then

$$M_o = \{o_i \in cd_j \mid a_x(c_x, cd_j)\} \quad (6)$$

That is, the predicate possible values are defined or the attribute values concrete area, or the relation values area.

When described rules are followed, are formed in form "subject-predicate-object" of the content semantic metadata elements. Restrictions on the items number in semantic metadata are not imposed.

Step 4: Similarity determining of user request and information resources semantic metadata triplets.

Analyzing the most popular of the strings similarity determining indicators, noting: none of them can't guarantee a good result when is changing the order of words in a sentence and using of multiple languages in a text string. In this regard, it is proposed not to use these metrics directly, and use the modified metrics that are worked based on the method of splitting words at the N-grams [13]. Using the so-called Shinhlinh method [14], which would create average N-grams, is modified of the formula for calculating the Severensen-Dicey coefficient which is a binary measure of strings similarity, using N-grams instead of characters:

$$p = \frac{2 \times (f_{NG}(X) \cap f_{NG}(Y))}{f_{NG}(X) + f_{NG}(Y)} \quad (7)$$

where: p – strings similarity factor: $0 \leq p \leq 1$,

X and Y – compared strings,

$f_{NG}(a)$ – function that determines the length of N-grams in the string a .

The usage of non-individual characters but word combinations allows to reduce the number of errors during of the current text string analysis.

The result that we get after the fourth step - is fully formed and stored in the system for further dynamic visual presentation combined data set that has the general structure and unique content.

4. CONCLUSIONS

The current methods of finding and receiving the data in dynamic data integration systems that are worked using "Mash-Up" technology have been researched and analyzed.

To improve of the quality indicators of Mashup system information storing and presentation result, the method of forming the united dynamic data set that has the general structure and unique content has been designed. The rules of forming the set of possible objects in triplets have been described. The procedure of the similarity determining of user request and information resources semantic metadata triplets has been characterized.

REFERENCES

- [1] KUSHNIRETSKA I., BERKO A.: *Application of Mash-Up Technology for Dynamic Integration of Semi-Structured Data*. Proceedings of the Sixth International Conference of Young Scientists CSE-2013, Lviv Polytechnic National University Publisher, Lviv, 2013, pp. 220–221.
- [2] ABITEBOUL S., GREENSHAN O., MILO T.: *Modeling the mashup space*. WIDM, 2008, pp. 87–94.
- [3] MAXIMILIEN E., WILKINSON H., DESAI N., TAI S.: *A domain-specific language for web apis and services mashups*. Proceedings of ICSOC'07, Berlin, Heidelberg, 2007, pp. 13–26.
- [4] FISHER T., BAKALOV F., NAUERZ A.: *An overview of current approaches to mashup generation*. Proceedings of the Fifth Conference Professional Knowledge Management: Experiences and Visions, Vol. 145, Solothurn, Switzerland, 2009, pp. 254–259.
- [5] *About IFTTT*. IFTTT Inc., 2015 [electronic resource] – Available on: <https://ifttt.com/wtf>
- [6] *Zapier*. Zapier Inc., 2015 [electronic resource]. – Available on: <https://zapier.com>
- [7] *About Pipes*. Yahoo! Inc., 2015, available on: <http://pipes.yahoo.com/pipes>
- [8] *Alerts. Follow for new interesting content on the Internet*. Google Inc., 2015 [electronic resource]. – Available on: <https://www.google.com/alerts>
- [9] BAADER F., CALVANESE D., MCGUINNESS D., NARDI D., PATEL-SCHNEIDER P.: *The Description Logic Handbook: Theory, Implementation and Applications*. Cambridge University Press, 2003, pp. 132–136.
- [10] LEVY A.: *Logic-Based Techniques in Data Integration*. Logic Based Artificial Intelligence. Kluwer Publishers, 2000, pp. 74–76.
- [11] RECCHIA G., LOUWERSE M.: *A Comparison of String Similarity Measures for Toponym Matching*. Proceedings of the First ACM SIGSPATIAL International Workshop on Computational Models of Place, 2013, pp. 54–62.

- [12] KUSHNIRETSKA I., KUSHNIRETSKA O., BERKO A.: *Determination of the structure and content of input information resources for Mashup system*. Technology Audit and Production Reserves, No. 6/3 (20), 2014, pp. 4–9.
- [13] BARASHEW D.: *The Similar Documents Search*. Computer science. Big Data'13, 2013 [electronic resource] – Available on: http://compscicenter.ru/sites/default/files/materials/2013_04_18_BigData_lecture_09.pdf.
- [14] GUDKOV V., GUDKOV E.: *N-grams in Linguistics*. Herald of Chelyabinsk State University, No. 24 (239), 2011, pp. 69–71.

*Rafał WOJSZCZYK**

THE MODEL AND FUNCTION OF QUALITY ASSESSMENT OF IMPLEMENTATION OF DESIGN PATTERNS

Abstract

One of the ways of providing high internal software quality (that is a source code) is using design patterns. The article aims at presenting a suggested model which enables one to assess the quality of implementation of design patterns. The model assumes verification of different aspects of the patterns and a numeric expression of the obtained results. The analysis of the obtained results may show the occurrence of certain problems which are difficult to be identified during code review or testing.

1. INTRODUCTION

Software is an intangible product, therefore, as a product, it shares some features with tangible products. The distinguishing feature in case of the two mentioned types of products is, among others, the fact that software is not subject to failure, wear or replacement. It seems that once software is produced it could be used forever. Unfortunately, as the reality has shown many times, software must undergo changes and be adjusted to this reality. An example of forced changes was implementation of new VAT rates in Poland in 2011. Not every software was prepared for this implementation without interfering into the source code or using the help of an IT specialist. The mentioned situation shows the importance of providing user-friendly enhancement or modification of software.

A similar example is adapting software to the clients' needs. Large ERP systems are very flexible, therefore, individual changes can be implemented by skilled specialists without interfering into the source code. In case of software produced by small companies, individual adjustment at the implementation stage is often impossible and this leads to changes in the source code. In order to avoid increased costs of these changes, it is necessary to implement suitable

* Department of Electronics and Computer Science, Koszalin University of Technology,
75-453 Koszalin, Poland, rafal.wojszczyk@tu.koszalin.pl

mechanisms which will allow for simple development, simultaneously, without disturbing the rest of software. One of the trusty solutions to the mentioned problems are design patterns.

Using design patterns is also very important at other stages of software development cycle. Another example encountered in reality is discontinuation of software development or a change of a development team responsible for certain software. Design patterns are recognized in the community of programming practitioners as a certain language of communication, therefore, the code of software containing design patterns will be more recognizable and under-standable than the one without the patterns.

2. SOFTWARE QUALITY AND DESIGN PATTERNS

Software quality is a very general notion, and a definition of the measure that would directly determine this value is impossible to be provided. When analysing different models of software quality and, especially, maintainability property included in the third part of ISO/IEC 9126 standard [6], one may conclude that the internal software quality (that is software code quality) is determined by the following features: easy development, easy modification, easy-to-understand code. Using design patterns in programming promotes positive values of the mentioned features (positive in a way that thanks to using design patterns, software is easier to be developed, etc.) [3]. Therefore, one can assume that the quality of design patterns implementation is one of the factors determining software quality.

The quality of design patterns implementation, similar to the general software quality, does not have any direct measure. Verification of correctness of the implemented design pattern' structure (in relation to the general template) is only one of the aspects of the process of assessment of design patterns. What is more, there are many other features which influence the implementation quality, e.g. distortion and bad smells, overlapping design patterns, typical errors, anti-patterns, and improper intention of using the patterns. Additionally, in the very structure of design patterns one can distinguish problems with verification, namely, considerable diversity of implementations and the occurrence of many variants of each pattern. There are yet many more problems connected with analysing implementation of design patterns - these were presented, among others, in [14] and [9]. These problems may lead to a misleading situation which maintains structural correctness of the pattern implementation but the use does not correspond with the intention or it does not serve its purpose.

The design patterns described in [3] are templates of ready-made mechanisms which can be used to solve typical problems occurring cyclically in object-oriented designing and programming. However, these are not ready-made solutions since using each pattern requires its proper implementation according to the software context.

The issues discussed in the research studies connected with design patterns very often concern the problem of searching for instances of design patterns in software [11, 10, 1]. The measure of the mentioned studies is presenting a number of instances of design patterns in software, which is insufficient to consider it as quality characteristics. What is more, there are studies conducted which mainly aim at showing structural correctness of implementation of the patterns [2], and those also provide too little information. Other attempts of numeric expression of design patterns were introduced in [7] and [8], and those consist in using existent object-oriented software metrics or author's metrics for software containing design patterns. Many years of research on metrics have provided the recommended values of metrics for the software generic case, which obtainment shows high quality. On the other hand, in [4] it was shown that the occurrence of design patterns does not have a positive influence on the results obtained from metrics, therefore, one should perform proper over-interpretation of metrics applied with the patterns, which was presented in [12].

3. ASSESSMENT MODEL

3.1. General assumptions

Figure 1 shows the general concept of the offered model. The first stage is acquiring software which will be subject to assessment of the quality of implementation of design patterns. The acquiring consists in converting software source code to a formal representation and, more specifically, to a proper data structure based on the assumptions of the object-oriented programming paradigm [13]. The analysis stage uses the acquired data solely in a form of the formal representation, and independent variables which are definition model, reference model and mechanisms used to verify the analysed software. The basic result is the vector of assessment obtained from assessment function for the analysed software. In the extended case, the result may include additional information concerning artefacts in the analysed software which show possible errors and change suggestions.

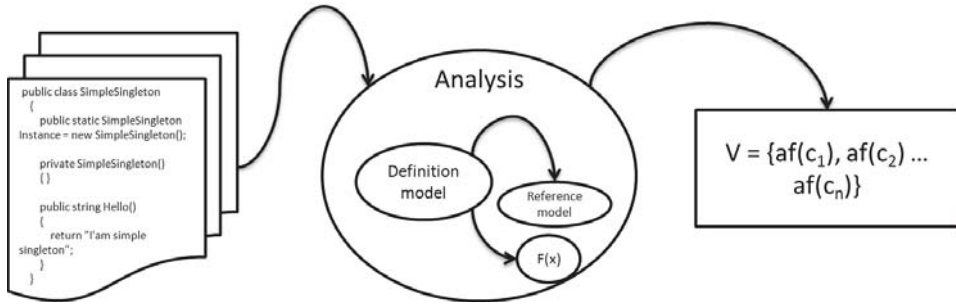


Fig. 1. General concept of the model [source: own study]

3.2. Definition model

The definition model is a hierarchical data structure describing the set of features required for verification in the analysed software. Figure 2 shows a symbolic hierarchy of the definition model. The root in the hierarchy is a specific design pattern since, for each design pattern, one should have a proper set of features. The first level of the hierarchy are general categories which group the features. The level of features concerns specific aspects of design patterns and the semantics of this level allows for defining relations "or", "and" as well as "include" between the features. Each feature may be broken down into elements describing the occurrence of detailed artefacts (that is, data types, components, instructions and others).

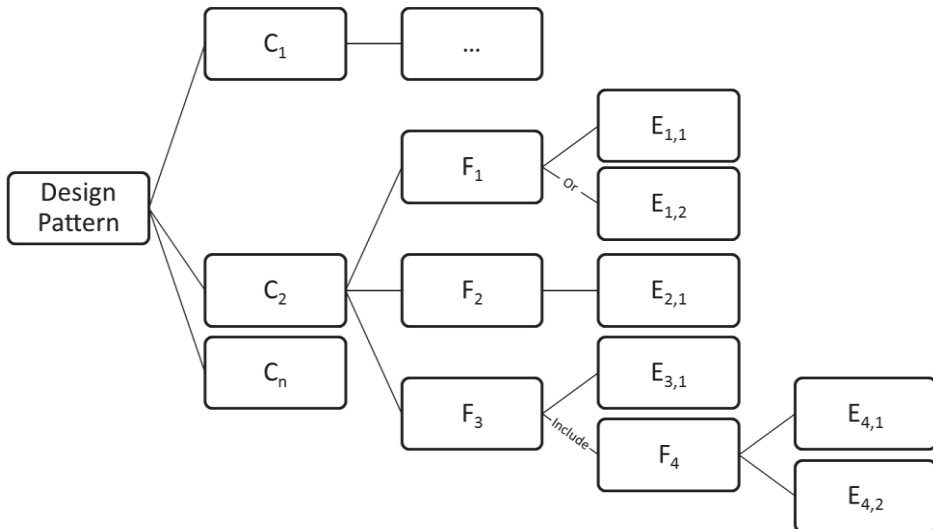


Fig. 2. Symbolic representation of the definition model [source: own study]

Below, one can find an exemplary definition of Singleton design pattern:

- structure – type: non-abstract class, no inheritance; instance: private, static field restored by public static property of type identical to the one of holding class, a suggested name is Instance; constructor: private,
- behavior – initialization: checking for object existence and creation at the first utilization,
- utilization – at least one shareholder and no more than 1/2 of the sum of classes and interfaces in the analysed software,
- connections to other patterns: abstract factory pattern,
- typical errors: other constructors than private ones are forbidden, there should be only one element functioning as an instance,
- other: a proper number of components (fields, methods) included in the pattern class, no inheritance from pattern class.

A principal role of the definition model is providing the description of the features to be verified; still, this model cannot conduct the mentioned verification. The definition model can only show one of the methods of the verification. Thanks to separating verification methods (in other words: operational definitions) from the definition model, increasing effectiveness of this verification is possible through selecting the most appropriate method. Additionally, this approach give great development possibilities in case previous verification methods turn out to be insufficient.

Features occurring in the definition model are diverse in relation to the analysed design pattern, similarly to the occurred categories. An exception is the structural correctness category which occur in every design pattern offered in [3], therefore, one of the verification methods was developed especially for this purpose. It is the verification of the analysed software in relation to the reference model. The reference model includes a general, abstract description of the structure of design patterns. It has been developed as an enhanced data structure based on the assumptions of object-oriented programming paradigm. It enables one to define different variants of the particular elements of design patterns with determining the level of their adjustment to the assumed ideal. Verification in relation to the reference model and other verification methods (e.g. making queries for the obtained software) may be used to verify any features; it is important for every method of verification to return proper result which then will be interpreted by the assessment function. The result is returned in the standardized form; it is the maximum in the interval scale; more specifically, it falls within the range from 0 to 1, where 1 stands for total occurrence of the analysed element. Table 1 presents a dataset of a single feature in the definition model. Each feature belongs to a category. Data included in the category is a name and assessment (identical to feature assessment).

Tab. 1. Dataset included in the feature [source: own study]

Data name	Description
Name	Verbal determination of the feature, understandable to a human.
Feature dependence	Indicating other feature and determining type of dependence related to it: and, or, include.
Assessment	Complex data structure used directly by the assessment function. It includes: weighting factor, activation threshold and recommended value, and, after verification is done, also the obtained result.
Multiplicity	Complex data structure containing information on the expected number of occurrences of a particular feature in the analysed software. It determines expected value in a form of the following ranges: from to, greater than or equal, less than, exactly.
Set of elements	Complex data structure which indicates the detailed artefacts and describes their possible variants. It includes: name identifying an element, assessment, that is, the structure identical to feature assessment, set of references to the reference model or other verification method, and negating the element, which means that all possibilities different from the indicated one are allowed.

3.3. Working principle

The above-described definition model has one more important role – it indicates subsequent stages of analysis. In the generic case, the process of analysis assumes the following: for each category one should verify desired features, while for each verified feature one should verify elements. Each stage of verification gives a partial result even if it does not satisfy the activation threshold. The final selection of results will be performed by the function initializing the assessment. Each verification stage is encumbered with checking for occurring restrictions, e.g. those connected with multiplicity (number) of occurring elements and performance of required verification operations.

The partial results are added to assessments of the corresponding elements (see Table 1 - Set of elements, Assessment). After conducting the required verifications, the function initializing assessment and assessment function are performed. Both functions are performed up the hierarchy, that is, initializing assessment is performed on the basis of results obtained from verification of elements; then, on the basis of a modified dataset, the assessment function is performed, which result is a result for the feature. Subsequently, this process is repeated for the features and, finally, for categories. The initializing functions checks the activation threshold and selection on the basis of relations between the features. If activation threshold is not satisfied then a particular element or feature does not occur; in special case, if activation threshold is 0, then occurrence of a particular element or feature is obligatory.

3.4. Execution

The offered model was executed as a prototype tool in Microsoft .NET technology. The used technology allowed for improvement of the process of acquisition of the analysed software, which consists in acquisition of object-oriented structure from CIL managed code using one of the methods of the reverse engineering, e.g. Mono.Cecil [13]. Using the managed code allowed for avoiding typical syntactic errors occurring in the source code and removed useless code fragments, e.g. unit test code, processor directives and other from the analysed software. An additional advantage resulting from the used technology is a possibility of applying the model to different programming languages compiled to the managed code. Formal data representation containing the acquired software was executed as a database in Microsoft SQL Server environment.

The analysis of the features of design patterns takes place according to the hierarchy described in the definition model which was also executed as a database in Microsoft SQL Server environment, similarly to the above-mentioned reference model. The worked out structure of the definition model allowed for accessible implementation of the major loop in which the subsequent features are iterated, and the analysis is conducted for each of those.

In case of "include" dependence, recurrence function is performed into the hierarchy; this is presented by listing 1. Possibility of verifying features by means of different mechanisms is a polymorphism property joining the definition model with verification mechanisms. After conducting the analysis, the results are recorded in an object-oriented data model creating the vector of assessment. An exemplary result of calculation is shown in Figure 3. After selecting a required category and then a feature (see Fig. 3), the software displays additional information resulting from the particular feature analysis: the partial results show attributes which occur for the analysed software (with level of adjustment in brackets); way of presentation of partial results is dependent on the used verification mechanism; in the symbolic representation of the analysed software, one can find hints concerning better solutions and error messages with improvement suggestions.

Listing 1. Major loop code

```
public void AnalyzeAll(Definition definition)
{
    foreach (Feature feature in definition.Features)
        AnalyzeRecursively(feature);
}

private void AnalyzeRecursively(Feature feature)
{
    Verify(feature.Elements);
    if (feature.KindOfDependence == KindOfDependence.Contain)
        AnalyzeRecursively(feature.DependenceFeature);
}
```

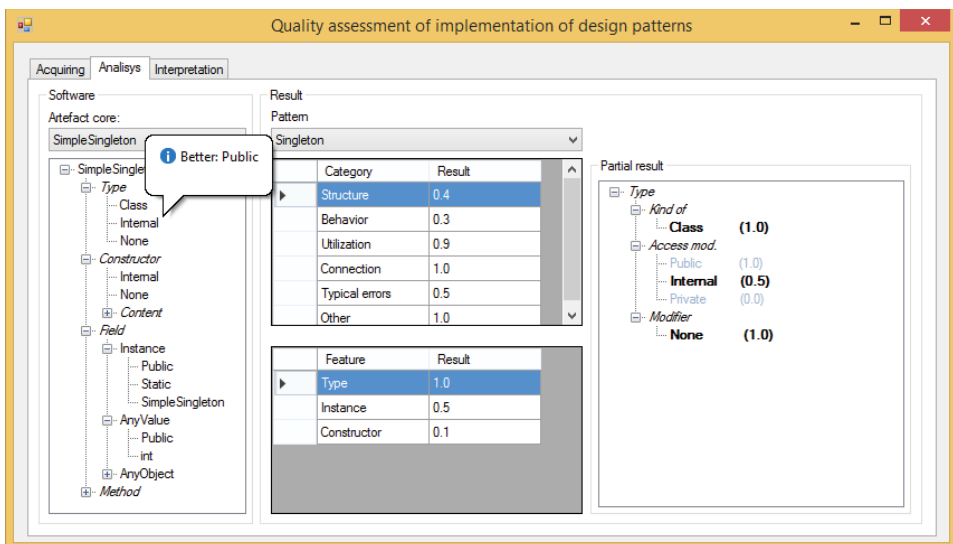


Fig. 3. Prototype execution [source: own study]

The main area of using the model is the whole process of software implementation where it can be used simultaneously with other tools taking care of generally understood quality, good practices and code cleanliness. It should be especially useful in companies or development teams which take care of providing high quality in the aspects of software maintenance and development. A very practical application would be integrating the model with programming environments to follow and assess changes connected with implemented design patterns on an ongoing basis. Finally, using the model in the opposite direction, that is, with at least semi-automatic implementation of the patterns in software on the basis of provided definitions is worth considering.

4. ASSESSMENT FUNCTION

The basic output of the offered model is vector of assessment of the following form

$$v = \{af(c_1), af(c_2), \dots, af(c_n)\}, \quad (1)$$

where: $af(c_1), af(c_2), \dots, af(c_n)$ is a result of assessment function for each subsequent category c_1, c_2, \dots, c_n . The quality of each of those categories is expressed by the following function

$$af(c_1) = af(f_1, f_2, \dots, f_n), \quad (2)$$

where: f_1, f_2, \dots, f_n is a set of features belonging to c_1 . Quality of the particular features in the generic case is determined by the following formula

$$af(f_1) = af(e_1, e_2, \dots, e_n), \quad (3)$$

where: e_1, e_2, \dots, e_n is a set of elements belonging to a particular feature. In a special case, where a particular feature is in "include" dependence in relation to other feature, the quality may be determined by the following function

$$af(f_1) = af(e_1, e_2, \dots, e_n, f_{1,1}, f_{1,2}, \dots, f_{1,n}), \quad (4)$$

where: e_1, e_2, \dots, e_n is once again a set of elements belonging to a particular feature, while $f_{1,1}, f_{1,2}, \dots, f_{1,n}$ are features of "include" dependence.

The basic formula describing the assessment function is

$$af(x) = \frac{\sum_{n=1}^n x_n * w_n}{\sum_{n=1}^n w_n}, \quad (5)$$

where: n – sum of arguments,
 x_n – particular argument (that is, element, feature, category),
 w_n – argument weighting factor.

One can notice that it is the weighted mean. Means (arithmetic or weighted) are often used in methods assessing software quality, e.g. in [5] or in recognized CK object-oriented software metrics [12]. The offered model allows a possibility of presenting other assessment function in case when the primary function turns

out to be insufficient. Weighting factors may initially be determined on the basis of expert knowledge or by method for valuation. What is more, the possibility of selecting proper weighting factors by means of artificial neural networks is taken into consideration.

5. INTERPRETATION OF THE RESULTS

An example of the obtained assessments rounded to decimal for three different occurrences of Singleton pattern was presented in Table 2.

Tab. 2. Exemplary results [source: own study]

	Case 1	Case 2	Case 3
Structure	1	0,4	1
Behavior	0,7	0,3	0,7
Utilization	0,6	0,9	0,1
Connection with other patterns	1	1	0
Typical errors	1	0,5	1
Other	1	1	0,3

Implementation of Singleton pattern in case 1 may be considered as of good quality. High structural correctness and connections with other patterns were maintained. The maximum assessment in typical errors category shows that these errors did not occur. Lower assessment in the behavior category results from a lack of reference synchronization to Singleton instance in multithread environments. Lower assessment in the utilization category results from too many references to instance which may indicate too high responsibility of the class with Singleton pattern.

Case no 2 showed low quality of implementation in the structural correctness, typical errors and behavior categories. Software containing such implementation may work improperly, e.g. there may be errors connected with data incompatibility or the so-called runtime errors. These errors should be detected already at the stage of software execution (testing stage) and repair cost should be on the budget (for error recovery) in the general production process. However, one should take into account that Singleton design pattern is considered to be one of the simplest patterns, therefore, the repair costs may be higher in case of other patterns.

Case no 3 showed low quality of implementation in categories of utilization, connections with other patterns and other features connected with too many components. The analysis of that case showed that the pattern was misused in connection to the intention of its use or it was improperly understood by the

implementing developer. The problems resulting from such implementation are difficult to be detected, both during code inspection and at testing stage. The most noticeable problems will occur during software maintenance and development. Repair costs are nearly unpredictable.

6. SUMMARY

The article briefly explains the influence of quality of implementation of design patterns on internal quality of software and presents selected research studies connected with verification of implementation of design patterns. It discusses the proposal of the model which allows for numeric expression of design patterns. The model consists of the hierarchical definition model and mechanisms responsible for verifying features of patterns. The result of the offered model is a vector of assessment which is obtained through the assessment function. The assessment function uses partial results acquired in the verification process and weighting factors from the definition model. The analysis of the vector of assessment and detailed results enables one to predict possible problems connected with software development and its improper working.

Further works related to the model include development of dependences of definitions occurring in the model (e.g. set of artefacts which satisfies requirements of a certain feature may be an input set for other feature), development of verification mechanisms and possibilities of analysing the results.

REFERENCES

- [1] BINUN A.: *High Accuracy Design Pattern Detection*. PhD Thesis, Rheinischen Friedrich Wilhelms Universität Bonn, 2012.
- [2] BLEWITT A.: *HEDGEHOG: Automatic Verification of Design Patterns in Java*. PhD Thesis, University of Edinburgh, 2006.
- [3] GAMMA E. et al.: *WZORCE PROJEKTOWE. Elementy oprogramowania wielokrotnego użytku*. Helion, Gliwice, 2010.
- [4] HERNANDEZ J. et al.: *Selection of Metrics for Predicting the Appropriate Application of design patterns*. 2nd Asian Conference on Pattern Languages of Programs, 2011.
- [5] HOŁODNIK-JANCZURA G.: *Badanie jakości produktu informatycznego metodą wartościowania*. Badania Operacyjne i Decyzje, Oficyna Wydawnicza Politechniki Wrocławskiej, ISSN 1230-1868, Wrocław 2007, pp. 55–69.
- [6] ISO/IEC TR 9126-3, Software Engineering – Part 3: Internal metrics, ISO/IEC 2003.
- [7] KHAER Md. A. et al.: *An Empirical Analysis of Software Systems for Measurement of Design Quality Level Based on Design Patterns*. Computer and information technology, IEEE, 2007.
- [8] MASUDA G., SAKAMOTO N., USHIJIMA K.: *Evaluation and Analysis of Applying Design Patterns*. IWPSE - International Workshop on Principles of Software Evolution, 1999.
- [9] RASOOL G.: *Customizable Feature based Design Pattern Recognition Integrating Multiple Techniques*. PhD Thesis, Technische Universität Ilmenau, Ilmenau 2010.

- [10] SINGH RAO R., GUPTA M.: *Design Pattern Detection by Greedy Algorithm Using Inexact Graph Matching*. International Journal Of Engineering And Computer Science, Vol. 2, No. 10, 2013, pp. 3658–3664.
- [11] TSANTALIS N. et al.: *Design Pattern Detection Using Similarity Scoring*. IEEE Transactions on Software Engineering, Vol. 32, No. 11, 2006, pp. 896–908.
- [12] WOJSZCZYK R.: *Zestawienie metryk oprogramowania obiektowego opartych na statycznej analizie kodu źródłowego*. Zarządzanie projektami i modelowanie procesów, Zeszyty Rady Naukowej Polskiego Towarzystwa Informatycznego, ISBN 978-83-7518-599-7, Warszawa 2013, pp. 95–107.
- [13] WOJSZCZYK R.: *Pozyskiwanie struktury obiektowej z kodu zarządzanego przy wykorzystaniu metod inżynierii odwrotnej*. Inżynieria oprogramowania: badania i praktyka, Zeszyty Rady Naukowej Polskiego Towarzystwa Informatycznego, ISBN 978-83-63919-15-3, Warszawa 2014, pp. 199–213.
- [14] WOJSZCZYK R.: *Koncepcja hybrydowej metody do oceny jakości zaimplementowanych wzorców projektowych*. Zeszyty Naukowe Wydziału Elektroniki i Informatyki nr 7, Wydawnictwo Uczelniane Politechniki Koszalińskiej, ISSN 1897-7421, Koszalin 2015, pp. 17–26.

Natalya SHAKHOVSKA*, Iryna SHVOROB**

THE METHOD FOR DETECTING PLAGIARISM IN A COLLECTION OF DOCUMENTS

Abstract

The development of the intelligent system for searching for plagiarism by combining two algorithms of searching fuzzy duplicate is considered in this article. This combining contributed to the high computational efficiency. Another advantage of the algorithm is its high efficiency when small-sized documents are compared. The practical use of the algorithm makes it possible to improve the quality of the detection of plagiarism. Also, this algorithm can be used in different systems text search.

1. INTRODUCTION

Nowadays, the Internet is the biggest source of information. Now, people can easily search, get access and browse the web to get the information they need. Just imagine how difficult it would be to do scientific research without the Internet and web space. Furthermore, due to the size and digital structure of the internet, it is easy to illegally use someone else's work now.

The problem of plagiarism has a direct relationship with the scientific community. The most common plagiarism is written text document which is formed by copying some or all parts of the original document, sometimes with some modifications. Identification of documents which were copied is stressful and time-consuming process to humans due to the large number of documents which have to be analyzed. The documents in digital format make the process of plagiarism quite simple, it means that such cases of plagiarism can be traced automatically.

* Lviv Polytechnic National University, Lviv, Ukraine, S. Bandery Str., 12, Lviv, 79013, natalya233@gmail.com

** Lviv Polytechnic National University, Lviv, Ukraine, S. Bandery Str., 12, Lviv, 79013, irka.shvorob@gmail.com

Plagiarism detection depends on many factors[5]. The first factor is the presentation of the document, which essentially covers the characteristics of the document as a preliminary step to compare [7, 8, 9]. These representations include model of identification tags, N-grams, probabilistic models, algorithms "scales" and others. Most of these representations work well in detecting verbatim plagiarism, but are vulnerable to identify complex patterns of plagiarism.

The second factor is a similarity and a measure of proximity, which is used to calculate the similarities or differences between sentences. Given the behavior of plagiarists, which usually includes insertion, deletion or substitution of words necessary to determine which activities are best for detection of plagiarism.

With the development of information systems the number of areas to identify plagiarism text only increased. This is the area of scientific papers, various publications in the field of journalism, fiction genres [13–15].

Currently, there are many methods and algorithms that can detect plagiarism of text objects. But over time, there are new challenges associated with the development of information systems. These tasks require more qualitative and more accurate detection of plagiarism in text.

Purpose of this paper is to improve the efficiency and quality of plagiarism detection in text objects by the use of the combined algorithm.

2. THE INTELLIGENT SYSTEM OF DETERMINE THE DEGREE OF RESEMBLANCE OF THE TEXTS

2.1. The algorithm development

Opt Freq algorithm implements the method of "optimal search frequency" and its used to search for similar documents in a wide range of applications, from web to clustering news. The gist of it is this. Instead of classical metrics $TF \cdot IDF$ a modified version of it is proposed. We introduce a heuristic concept of "optimal frequency" for the word "equal" $\ln\left(\frac{10}{1000000}\right) = 11.5$ which means "the optimal" entering of word in 10 documents from 1000000. If the real value of IDF is less than "optimal", then it slightly (by law parabola) rises to $IDF_{opt} = \sqrt{\frac{IDF}{11.5}}$, and if it is greater it significantly (as hyperbole) reduces to

$$IDF_{opt} = \sqrt{\frac{11.5}{IDF}} \quad (1)$$

For the collection the dictionary is created. This dictionary puts every word in accordance with the number of documents in which this word occurs at least once (df). Then the frequency dictionary for document is built and the “weight” wt of each word is calculated by the formula:

$$wt = TF * IDF_{opt} , \quad (2)$$

where:

$$TF = 0.5 + 0.5 * \frac{tf}{tf_{max}} , \quad (3)$$

$$IDF == \log\left(\frac{df}{N}\right), \quad (4)$$

$$IDF_{opt} = \begin{cases} \sqrt{\frac{IDF}{11.5}}, & IDF < 11.5 \\ \frac{11.5}{IDF}, & IDF \geq 11.5 \end{cases} \quad (5)$$

tf (term frequency) is the ratio of occurrences of a word to the total number of words of the document. Thus, the estimated importance of words within a single document:

$$tf = \frac{n_i}{\sum_k n_k} \quad (6)$$

where n_i is the number using the word in a document, and the denominator – the total number of words in this document.

df (inverse document frequency) is inversion frequency with which a certain word is found in the documents collection. Consideration df reduces weight widely used words:

$$df = \log \frac{|T|}{|T_i \supset a_i|} \quad (7)$$

where $|T|$ is count of text documents in collection; $|T_i \supset a_i|$ is count of text documents, where word a_i occurs (where $n_i \neq 0$).

Then the 6 words with the largest values of wt are selected and concatenated in alphabetical order into the string. The check sum of the resulting line is calculated as the signature of document [3, 6].

Also, it is very important, where part of text is arisen [10, 11].

First of all, we introduce the concept of weight sentence.

$$Location = \frac{1}{n*m} \quad (8)$$

where $n = \overline{1..3}, m = \overline{1..3}$ – the place calls to the main part and paragraph respectively. Begin and end of text or paragraph estimated value of 1, the middle is as 3. Coefficient key phrase is determined by entering the sentence U of elements of a set of significant sentences from A membership function:

$$Cuephrase = \mu_A(U) \quad (9)$$

$A = \{\text{«Conclusion»}, \text{«In the end»}, \text{«By the way»} \dots\}$.

Index of statistical significance is formed on the basis of visiting sentence key-words specified by the author of the article:

$$Statterm = \mu_K(U) \quad (10)$$

The value added is defined as the presence of terms related words sentences that appear in the article's headline to the total number of words in a sentence (words) except for words whose length is less than 3 characters:

$$Addterm = \frac{word}{words} \quad (11)$$

The weight of text block U is:

$$Weight(U) = Location(U) + Cuephrase(U) + Statterm(U) + Addterm(U) \quad (12)$$

So after being allowed to study all the documents necessary to accomplish the following: to exclude a statement that its content has hit the consolidated data repository and perform the final sorting sentences. For the task of bringing to the final ranking factor "information novelty" use the following method:

- Let we have two sets of sentences $B = \emptyset$ and $A = \{A_i | i = 1, 2, \dots, N\}$, N is count of sentences in text. For every sentence A_i the usefulness $P(i)$ q_i :

$$P(i)_i = q_i, \quad i = 1, 2, \dots, N \quad (13)$$

- The sentences from set A sort Descending $P(i)_i$
- If A_i has the biggest $P(i)_i$, we take it in B. The usefulness for sentences in A set s

$$P(i) = \frac{P(i)}{kq_i} \quad (14)$$

where $k > 0$ – factor clipping similar sentences.

- Is A empty? If NOT, go to 1.

The next problem is information estimating from different sources [15–16]. For semi-structured data type text file with a known format – dictionary data types defined formatting released the text of the formatting, copying its contents:

$$object \rightarrow Find\left(\pi_{firmatttype}\left(\sigma_{object}(Dic)\right)\right) \quad (15)$$

```
foreach object
Selection
. ParagraphFormat.Alignment = Left (1, formattype)
. Font.type = Mid (formattype, 3, 1)
. Font.Caps = Right (formattype, 1)
InStr (1,. Text, Right (formattype, 2);
Copy.Selection
```

2.2. The proposed algorithm

The proposed algorithm can be divided into the following stages.

1. The construction of the dictionary of words.

The vocabulary is created throughout the collection. Each word is associated with a number of documents that it occurs at least once (df) and the average length of the document is determined (dl_avg).

2. The construction of the frequency dictionary.

The frequency dictionary to document is constructed and to each word its "weight» wt on a formula Okapi BM25 with parameters k = 2 and b = 0.75 is calculated.

3. The construction of signatures.

11 I-Match signatures created for each document. The basic idea of this approach is to calculate the daktilohra I-Match for presentation of the documents content. For this purpose, initially for a collection of documents created dictionary L, which include words with average values IDF, because these words provide are usually more accurate results in identifying fuzzy duplicates. Words with large and small values IDF rejected.

Then, for each document set of different words U is formed belonging to it, and determined the intersection of U and L. If the size of the crossing of vocabulary over some minimum threshold (determined experimentally), the list of words that are included into intersection are ordered, and it is calculated I-Match signature (hash function SHA1). In addition to the main dictionary L created K various dictionaries L1-LK, obtained by accidental deletion from the initial dictionary some fixed small part of words p, part of about 30–35% of the original amount L.

For each document, instead of one calculated $(K + 1)$ I-Match signature by the described above algorithm, the document that is represented as a vector of dimension $(K+1)$ and two documents are considered duplicates if they match at least one of the coordinates.

If the document undergoes small changes (order n words), then chances are that there are at least one of K additional signature remains unchanged, is:

$$1 - (1-pn)^K (*)$$

Indeed, the likelihood that change will not affect any other dictionary is pn - the probability that all changes fall into a remote part of the original dictionary. Then $(1-pn)$ - the probability that the signature will change, and $(1-pn)^K$ - the probability that all signatures will change (because additional dictionaries formed independently), so $(*)$ – is the desired probability.

Recommended values of parameters that are well proved in practice $p = 0.33$ and $K = 10$.

4. The find of duplicate.

Documents which have at least one coincidence of signatures considered as duplicate (There is a conflict of hash codes).

3. THE SYSTEM ARCHITECTURE

To build an information system model is used CASE-tool AllFusion Erwin Data Modeler , which enables model based infological model of information system build its datalogical model and create a database in any database management system. The development of the summarization system provides in the notation IDEF1X.

During the implementation of systems analysis for this area following charts were developed[4]:

1. IDEF0-diagram for subtasks of the main business process (figure 1);
2. IDEF3-diagram for the block "Choosing the algorithm of working with words" (figure 2).

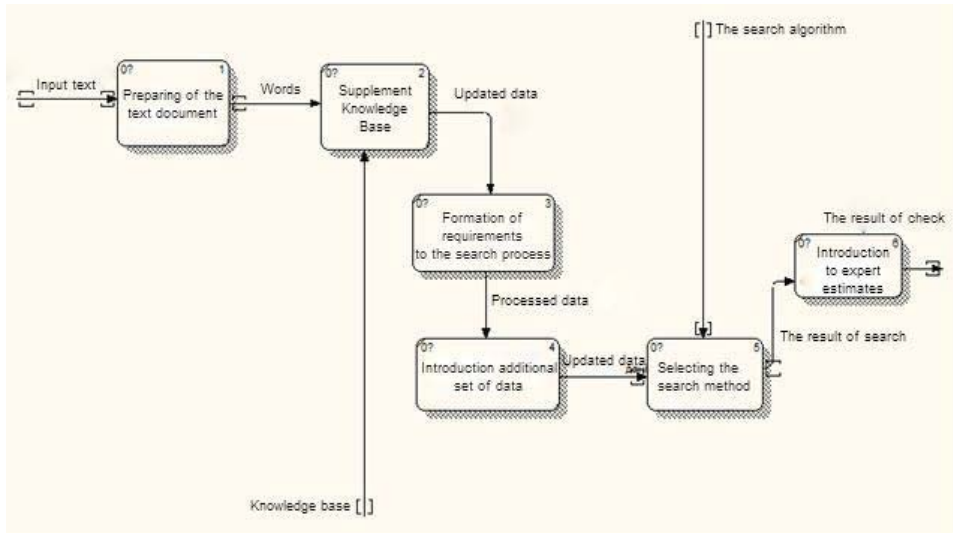


Fig. 1. IDEF0-diagram for subtasks of the main business process [own study]

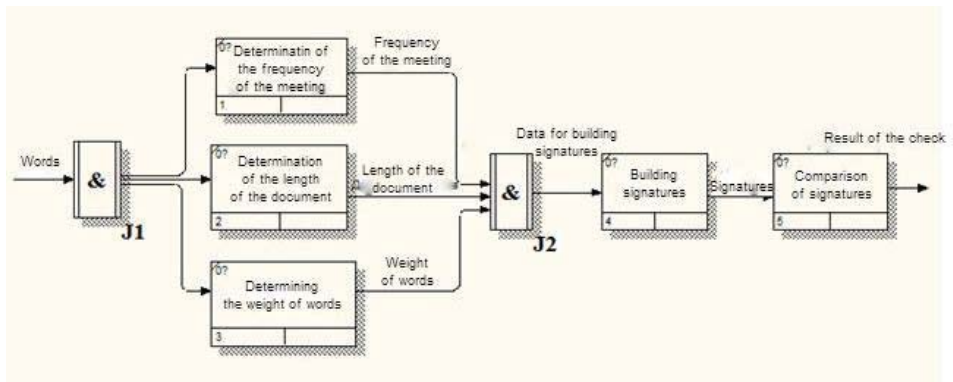


Fig. 2. IDEF3-diagram for the block "Choosing the algorithm of working with words"[own study]

The work "Checking for plagiarism in the text" is divided into 6 works: "Preparing of the text document", "Supplement Knowledge Base", "Formation of requirements to the search process", "Introduction additional set of data", "Selecting the search method", "Introduction to expert estimates". These works are carried out in the system sequentially, one after another. A text document which gets into the system due to user actions is applied to the input to the work "Preparation of the text document". Text information namely data entered in the system after this work is the result and therefore the input information for the work "Supplement knowledge base". This information is converted into data format suitable for the system in which they are ready for further processing.

Checking words of text is carried out as a result of the "Supplement Knowledge Base". The result of this work is a set of sentences which will be applied to the input of the "Formation of requirements to the search process" for further processing and the input of "Introduction additional set of data". These works are carried out the text processing. Words and formed signatures are the result of the work "The introduction additional set of data". They apply to the input of "Selecting the search method" and then searching for plagiarism is carried out. The next work "Introduction to expert estimates" provides the end result – a numeric value of the searching for plagiarism.

The IDEF3-diagram for the block "Choice of algorithm with the words" is consists of such units of work: "Determination of the frequency of the meeting" (determines the number of meeting of words in the text, returns the number of meeting of words in the document), "Determination of the length of the document" (determines the length of the document), "Determining the weight of words" (the data obtained in previous studies and knowledge base are used and keywords of the text are assigned of weight), "Building signatures" (connecting words into signatures), "Comparison of signatures" (checking of signatures, a collision of hash codes takes place).

The proposed system has a large number of works. These works are different, and usually independent. So, for greater flexibility they should be divided on the modules.

The system consists of the following modules:

- the database and knowledge;
- the subsystem of integration and information gathering;
- the subsystem of analysis.

Database and knowledge is designed to collect structured data and meta data about used text. It is a central part of the program because it is used by all the other parts

Whereas the system provides integration with other systems and sources of research documents, it must have functionality that would allow it to read the files and render its data to other systems. The subsystem of integration and data collection corresponds for this.

The analysis is carried out based on the downloaded document. The subsystem of analysis implements the proposed document analysis algorithm.

Input data are:

- 1) file with the text for which the test will be performed. File format .doc or .docx;
- 2) file with the text, which will be carried out the test input document. File format: .doc or .docx.

The result of the algorithm is a numerical value of the degree of the resemblance of the texts. This refers to a real number within a $[0, 1]$. If the result returns 0, it means that the texts are different, if 1 – text completely identical.

There also conclusion will be displayed: if the numerical value of the degree of resemblance of the texts of more than 0.3, the selected text – plagiarism, if less than 0.3, it is not plagiarism.

Figures 3a and 3b shows the software implementation of the developed system and the results of its implementation.

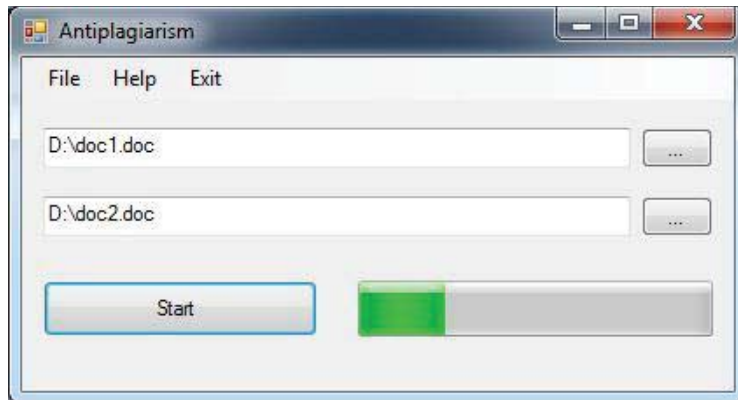


Fig. 3a. The test example of the program [own study]

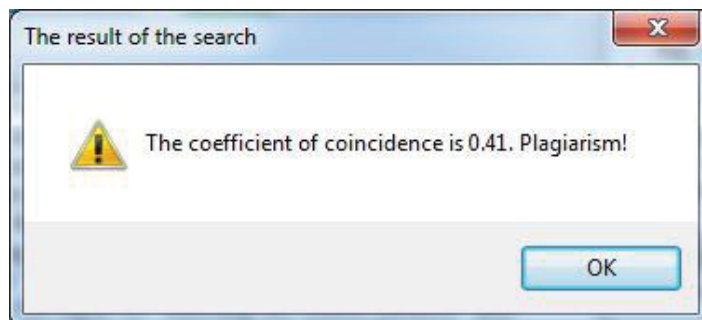


Fig. 3b. The results of the program [own study]

Before checking for plagiarism the system must makes its preliminary treatment.

This stage applies to all requested documents, as well as primary documents. There are four steps at this stage:

- remove stop words;
- fragmentation of the text;
- tokenization of the text;
- selection of roots of words.

4. CONCLUSION

The problem of plagiarism has been established and discussed. Two algorithms for finding fuzzy duplicates were considered and incorporated. The verification of the considered algorithms and combined algorithm was made. The received data are shown in Table 1. It is worth noting, as a result of combining it has improved its performance and result of searching of fuzzy duplicates.

System analysis for the intelligent system of determines the degree of resemblance of the texts was carried out and two charts were developed.

Basic steps for preprocessing the text were identified.

Tab. 1. The results of verification of algorithms [own study]

	Lex Rand Algorithm	Algorithm Opt Freq	Combined Algorithm
The accuracy of searching of plagiarism	39	59	61
A validation (sec.)	35	41	39

Obviously, the algorithm is not perfect in solving the problem of determining fuzzy duplicates. In order to improve options for combining multiple algorithms. For example, using the method of "descriptive words" can determine what class includes documents are scanned as each generated vector uniquely identifies this class. Then identify duplicates in a particular class of documents, signatures using methods based on the analysis of special characters. In this case, the possible increase effectiveness duplicate determination in a particular class of documents.

Duplication of texts in information flows is not always a negative phenomenon in terms of the user who uses the Internet for business purposes. An example of such an exception, for example, ranking brand when republication counts the number of press releases. Also you can use a number of overlapping signs "measure of importance" of a message and more.

REFERENCES

- [1] PARK S.-T., PENNOCK D., LEE GILES C., KROVETZ R.: *Analysis of Lexical Signatures for Finding Lost or Related Documents*. Finland, 2002, p. 8.
- [2] NIKOL'SKIJ J., PASICHNIK V., SHHERBINA J.: *Sistemi shchuchnogo intelektu*. Vidavnictvo Magnolija-2006, L'viv 2010, p. 279.
- [3] ZELENKOV J., SEGALOVICH I.: *Sravnitel'nyj analiz metodov opredelenija nechetkih dublikatov dlja Web-dokumentov*. Devjataja konferencija KSB, 2007.
- [4] KATRENKO A.: *Systemnyj analiz: pidruchnyk z gryfom MON*. Magnolija-2006, L'viv 2009, p. 352.
- [5] MAURER H., KAPPE F., ZAKA B.: *Plagiarism – A Survey*. Journal of Universal Computer Sciences, Vol. 12, No. 8, 2006, pp. 1050–1084.
- [6] The Open Archives Initiative Protocol for Metadata Harvesting Protocol Version 2.0 of 2002-06-14. <http://www.openarchives.org/OAI/2.0/openarchivesprotocol.htm>
- [7] HRYTSENKO V.: *Ynformatsionnye tekhnolohy: tendentsyy, puty razvytyia*. Upravliaiushchye systemy u mashyny, No. 5, 2001, pp. 3–20.
- [8] KRAIOVSKYI V.: *Osnovni pidkhody do rozroblennia prohramnoho kompleksu avtomatychnoho referuvannia tekstovykh dokumentiv* [in:] Kraiovskiy V. I., Lytvyn V. V., Shakhovska N. B., Zbirnyk naukovykh prats NAN Ukrainy, Instytut problem modeliuvannia v enerhetytsi, No. 51, Kyiv, 2009, pp. 178–186.
- [9] PARK S.-T., PENNOCK D., LEE GILES C., KROVETZ R.: *Analysis of Lexical Signatures for Finding Lost or Related Documents*. Finland 2002, p. 8.
- [10] KOŁCZ A., CHOWDHURY A., ALSPECTOR J.: *Improved Robustness of Signature-Based Near-Replica Detection via Lexicon Randomization*. KDD, 2004.
- [11] BRODER A.: *Identifying and Filtering Near-Duplicate Documents*. COM'00, Proceedings of the 11th Annual Symposium on Combinatorial Pattern Matching, 2000, pp. 1–10.
- [12] BERSON T.: *Differential Cryptanalysis Mod 232 with Applications to MD5*. EUROCRYPT. <http://dl.acm.org/citation.cfm?id=1754956>.
- [13] HAHN U., MANI I.: *The Challenges of Automatic Summarization*. Computer, Vol.33, No. 11, 2000, pp. 29–36.
- [14] Document Understanding Conferences (DUC): Web site, 2008. <http://duc.nist.gov>. 15.10.2011.
- [15] ALYIGULIEV R.: *Avtomaticheskoe referirovanie dokumentov s izvlecheniem informativnykh predlozheniy*. Vyichislitelnyie tehnologii, Vol. 12, No. 5, 2007, pp. 5–15.
- [16] YANG C., WANG F.: *Fractal Summarization for Mobile Devices to Access Large Documents on the Web*. Proc. of the WWW2003, May 20-24, 2003, Budapest, pp. 26–31.

Wojciech GRODZKI*, Andrzej ŁUKASZEWICZ**,
Kacper LEŚNIEWSKI***

MODELLING OF UAV'S COMPOSITE STRUCTURES AND PREDICTION OF SAFETY FACTOR

Abstract

The paper presents possibilities of composite materials modeling using SolidWorks environment on example of Unmanned Aerial Vehicle wing structure. Mechanical properties of composite materials used in UAV's and process of modeling such structures in SolidWorks are described. The research problem is CFD and strenght analysis of considered structure in SolidWorks Flow Simulation. Different displacement, stresses and safety factors values were obtained for analysed types of loads. The presented approach was used to develop aircraft wing for Air Challenge 2015 competition.

1. INTRODUCTION

Unmanned Aerial Vehicles (UAV) gained high popularity in the civil applications. They are widely used in various fields of life like: polar region monitoring, research, studying atmospheric pollution, counting animal populations etc.

Today's engineers task relies on design and manufacture products at reasonable prices and short time. It is possible due to the development of computer systems and wide availability of modern composite materials. Polymer matrix composites are commonly used in today's industry due to low fabrication cost and satisfactory mechanical properties. They mostly occur in form of symmetrical and asymmetrical laminates. Low weight, high strength, high stiffness, corrosion resistance and vibration damping ability features polymer

* Faculty of Mechanical Engineering, Białystok University of Technology, Poland,
wojtekgr@interia.pl

** Faculty of Mechanical Engineering, Białystok University of Technology, Poland,
a.lukaszewicz@pb.edu.pl

*** Faculty of Mechanical Engineering, Białystok University of Technology, Poland,
kacper.a.lesniewski@gmail.pl

matrix composites reinforced with continuous fibre. Thanks to high strength-to-weight and stiffness-to-weight ratios they are used for lightweight structures in automotive, sport industries and aviation (as parts of airframes) [1].

Nowadays technical universities become serious competitors against specialized companies in developing UAV designs. It is possible due to dynamic progress in computer technology and electronics. Modern computer-aided design tools are widely available and used by both industry and universities for engineers education [2]. Using the CAX tools meaningfully decreases price of the new product introduction. Results obtained from numerical analysis for safety reasons need to be compared with experimental data. CAX environment gives the possibility of designing 3D model (CAD), then performing series of numerical simulations like strength and flow analysis (CAE), technology development (CAM) and finally creation of 2D documentation. Computer aided design however, is only a tool in the hands of engineers, therefore it is important to choose or elaborate suitable strategies for creation a virtual model of real objects [3]. We can distinguish types of 3D modelling such as solid, surface, hybrid (surface-solid) and multibody. Usage of these techniques depends on the complexity and purpose of project. Computer systems like SolidWorks provide tools for modelling structure of composite materials and analysis of their mechanical properties. The software allows to define new types of material by describing their material properties (isotropic, orthotropic materials).

2. MODELLING OF COMPOSITE MATERIALS IN SOLIDWORKS

SolidWorks Simulation provides tools for modelling and analysing structures based on composite materials (the mentioned module requires surface model of designed component). Software environment allows to create surface model (3D CAD module) and use designed geometry in Simulation for modelling the composite structure. Another solution relies on creating preliminary solid model (3D CAD module) and then generating shell elements in SolidWorks Simulation module based on given geometry. This function gain importance in further analysis (strength) based on CFD resultant loads (Flow simulation requires 3D solid model for analysis).

Abovementioned environment gave possibilities to analyse composite laminates up to 50 different layers. SolidWorks assumes perfect bond between plies. Most common types of composites are (Fig. 1):

- Symmetric laminates defined by symmetric arrangement of ply materials, orientations, thicknesses, fibre orientation about the mid plane.
- Asymmetric laminate characterized by no symmetry layers, orientations and properties in reference to the mid plane.

- Sandwich composites represents case of symmetric laminates with mid plane layer defined by greater thickness and lower mechanical properties compared to other plies. Structure suitable when higher resistance to bending load is required.



Fig. 1. Types of composites (right-symmetrical, asymmetrical, sandwich composite)

SolidWorks composite property manager allows to define required thicknesses, orientations, numbers of layers and ply materials for designed structure. Graphic window of this tool shows laminate and global coordinate system with direction of each individual layers. It is very important to notice that coordinate system of composite is not the same as model or assembly. Defined composite structures position can be controlled relative to its surface. Composite property manager allows to position designed stack of plies middle surface, top surface, bottom surface and at specific ratio (position defined by offset value that is fraction from total thickness of structure, measured from the mid-surface to the reference surface). Fig. 2 presents symmetrical laminate consisting of 2 plies of 0.25 mm thicknesses, rotated with respect to material coordinate system by 45° based on carbon-fabric/epoxy material and surface mapping (software provide also planar mapping used to project a common 0° reference to a group of surfaces).

Important step in the modelling of composite structures is defining materials properties. SolidWorks environment for complete characterization of the orthotropic materials (composites used in analysis) requires describing properties such as elastic modulus E , Poisson ratio ν , shear modulus G , shear strength, tensile strength and compressive strength in specific directions [4].

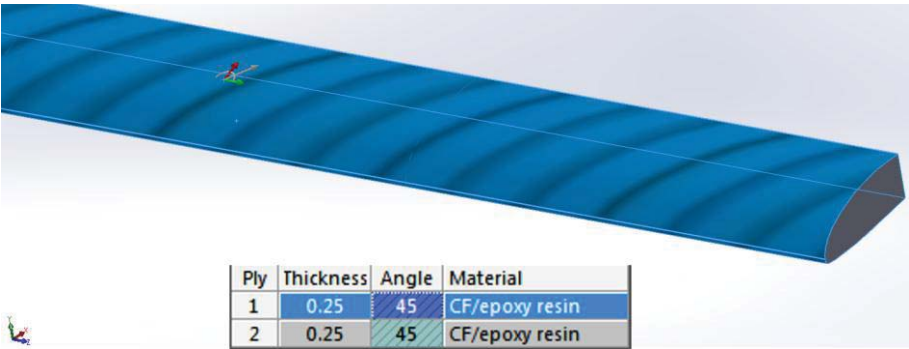


Fig. 2. Orientation, thickness, layer, material and coordinate system of composite

Results obtained from composite materials strength analysis apart from the standard strength analysis plots are: maximum stress across all plies, stress along the ply orientation direction and transverse direction to ply angle, stress on top or bottom face of each ply, interlaminar shear stress at the contact between two different plies. It should be noticed that strain fields and displacements are continuous through thickness of considered composite, while stress fields are discontinuous due to different orientations and material properties across plies.

SolidWorks contains three theories available for calculating laminate failure criteria: Tsai-Hill Failure Criterion, Tsai-Wu Failure Criterion, Maximum Stress Criterion. Based on one of these criteria we can generate safety factor plots (FOS) for our designed structure. Failure of composites occurs in several steps. When stress in the first ply or group of plies is high enough, it fails. This point of failure is called the first ply failure (FPF) beyond which considered laminate still carry the load. For safety reasons laminates should not experience stress high enough to cause FPF [5].

3. DESIGN AND NUMERICAL ANALYSIS OF COMPOSITE WING STRUCTURE

Design of unmanned aerial vehicles wing structure is multistage task that requires: defining UAV's purpose, aerofoil selection, geometrical calculations, structural design, materials selection, numerical analysis and elaboration of technology. Preliminary research of UAV wing structure concerns calculation of geometrical main dimensions (wing span, root and tip chords, twists, dihedrals and aerofoil distribution) based on the project assumptions and selected aerofoils [6,7]. Figure 1 presents results of initial study performed in XFLR5 environment - designed geometry (aerofoil cord 400mm, wing span 3000 mm) characterize by 310 N lifting force obtained at air velocity equal to 20 m/s (Fig. 3).

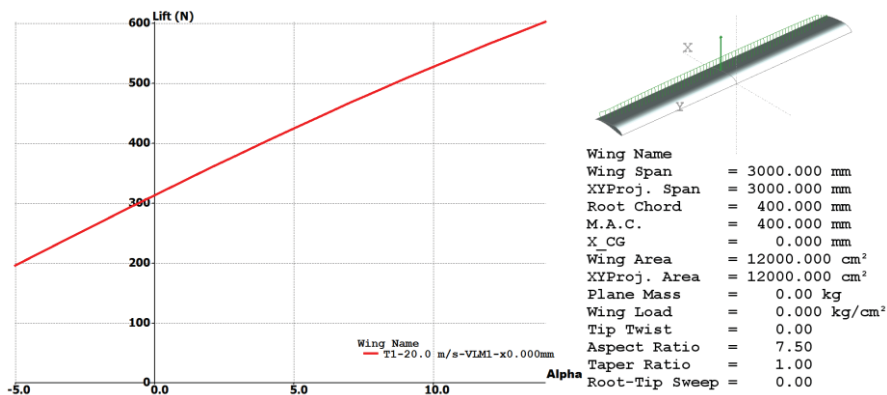


Fig. 3. Designed wing geometry in XFLR 5

XFLR 5 is a free software that allows numerical analysis of aerofoils, wings and airframes. In simulation this software uses a non-linear lifting line method, vortex Lattice Method (VLM) and 3D panel method. The results are lift and drag coefficients in case of aerofoils and lift and drag forces in case of wings and airframe according to defined geometry. Abovementioned software operates at Reynolds numbers [8].

The knowledge about types of loads acting on aircraft wing is essential during design process of new construction. In case of UAV wing structures most important are bending loads and torsion loads derived from acting lift force. Based on CFD simulation we are able to determine their types and values. Figure 4 shows flow simulation analysis of considered wing along with resultant lift force. On the basis of CFD simulation types and values of loads were obtained: average pressure acting on wing 275 Pa, concentrated force placed in geometric centre of wing equal to maximum lift 327 N.

Differences between values of lift force from XFLR 5 environment (315 N) and SW Simulation (327 N) results from type of research. In order to improve Flow simulation calculation time, analysis was simplified to 2D issue which does not take into account the decrease in lift force on the ends of the wings resulting from mixing of air masses with a high and low pressure.

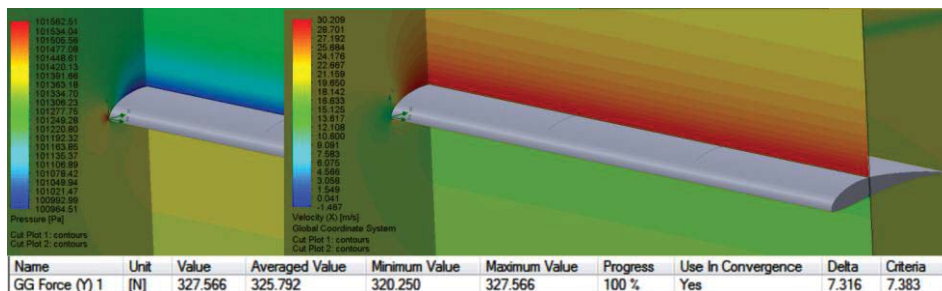


Fig. 4. CFD analysis of considered wing geometry (2D), results – lift force

Designed wing geometry (3D solid model) analysed in Flow simulation in order to achieve values and types of loads for further studies (modelling composite structure, strength analysis) requires geometry type change to surface model. Main bearing element of considering wing construction is caisson (part of an aircraft wing design to transferring bending and torsional loads, resulting from the impact of aerodynamic forces. Located on the front portion of the aerofoil, taking form of a closed thin-walled section). Design of composite structure (Simulation module) for further study is based on surface model of wing caisson filled with carbon fibre ribs. Created wing structure represented by symmetrical laminate consisting of 2 plies of 0.25 mm thicknesses, rotated with respect to material coordinate system by 45° based on carbon-fibre/epoxy material is shown at Figure 2.

Strength analysis of designed wing (SW Simulation) will be carried out for two cases of boundary conditions resulting from different types of loads: average pressure acting on wing (763 Pa) derived from acting lift force and concentrated force (315 N) placed in centre of wing simulating conditions of 2,5g overload (safety reasons) (Fig. 5). For analysis it was necessary to create finite element mesh equally distributed for the whole model (global size of the element 3.9 mm, tolerance 0.18). Figure 5 presents discussed boundary conditions.

For the numerical analysis iterative solver FFE Plus were applied. This method uses approximate techniques to calculate the solution and repeat the process until the difference between two consecutive solutions is significantly small or does not exceed set error limits.

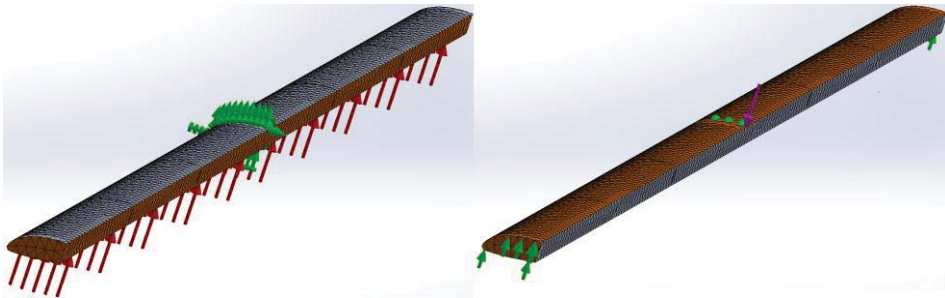


Fig. 5. Boundary conditions of two types of applied loads

Description of the research results will begin with displacement graph analysis (Fig. 6). It shows the resultant displacement (it is possible to check displacement in requested axis) of the analysed element under the influence of applied forces and geometrical restraints. Static test conditions caused three times larger wing deflection (124.2 mm) compared to load simulating flight conditions (40 mm). Higher structure stiffness (lower deflection) may be obtained by increasing thickness of the layers, adding number or changing rotation of plies.

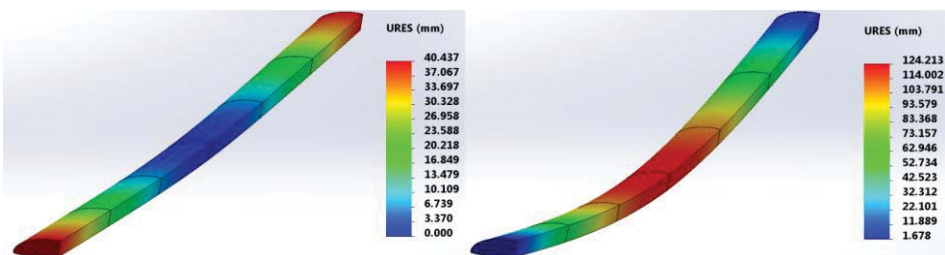


Fig. 6. Displacement of carbon fabric lamina (left-concentrated force load, right pressure load)

The following plot presents distribution of highest values and types of stress occurring in researched structure. Figure below (Fig. 7) shows values and distribution of tensile and compressive stresses appearing on carbon fibre layers of examined wing. It is possible to study stress values on each individual layer. Concentrated force placed in centre of the wing, simulating conditions of static test as predicted caused higher stress value equal to 260 MPa of tensile and 236 MPa of compressive stress compared to 104 MPa of tensile and 85 MPa of compressive stress in case of load simulated by pressure. Tensile stress appeared mostly on the lower part of construction while compressive stress occurred on the upper part of wing.

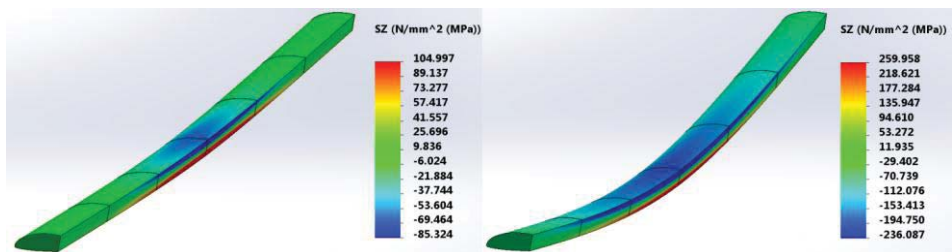


Fig. 7. Tensile and compressive stresses in considered structure

Last evaluated graph present distribution of factor of safety (FOS) over wing structure according to Tsai-Wu failure criteria (Fig. 8). This criterion considers the total strain energy (distortion and dilatation energy) for predicting failure. It is more conservative than the Tsai-Hill failure criterion because it distinguishes compressive and tensile failure strengths. The Tsai-Wu failure criterion cannot predict different failure modes such as fibre failure, matrix failure etc. Lowest FOS value across all plies equal to 1.6992 was obtained in case of study simulating static test. Factor of safety larger than 1.0 indicates that the laminate is safe from failure.

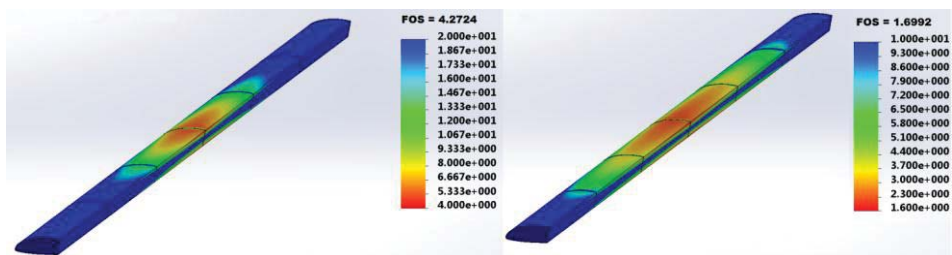


Fig. 8. Factor of safety (Tsai-Wu criterion) of carbon fibre lamina according to two type of load

4. CONCLUSION

Conducted analyses allowed to determine structure behaviour and stress occurring in respect to given geometry and used materials. Gathered information helps designer to create structure satisfying the requirements set out in the project. Results obtained from composite materials strength analysis gave possibility to evaluate additionally maximum stress across all plies, stress along the ply orientation and transverse direction to layer angle also stress on top or bottom face of each ply. For safety reasons laminates should not experience stress high enough to cause FPF (first ply failure). Higher structure strength (lower deflection) maintaining the same mass of the element may be obtained by changing rotation of composite layers. Analysed structure of the wings based on the carbon laminate satisfies Tsai-Wu failure criteria obtaining lowest FOS value across all plies equal to 1.6992. Using the CAx tools meaningfully decrease price of the new product introduction. Results obtained from numerical analysis for safety reasons should be compared with experimental data. Presented approach was used to develop aircraft wing structure based on carbon fabric epoxy resin laminate (Fig. 9).

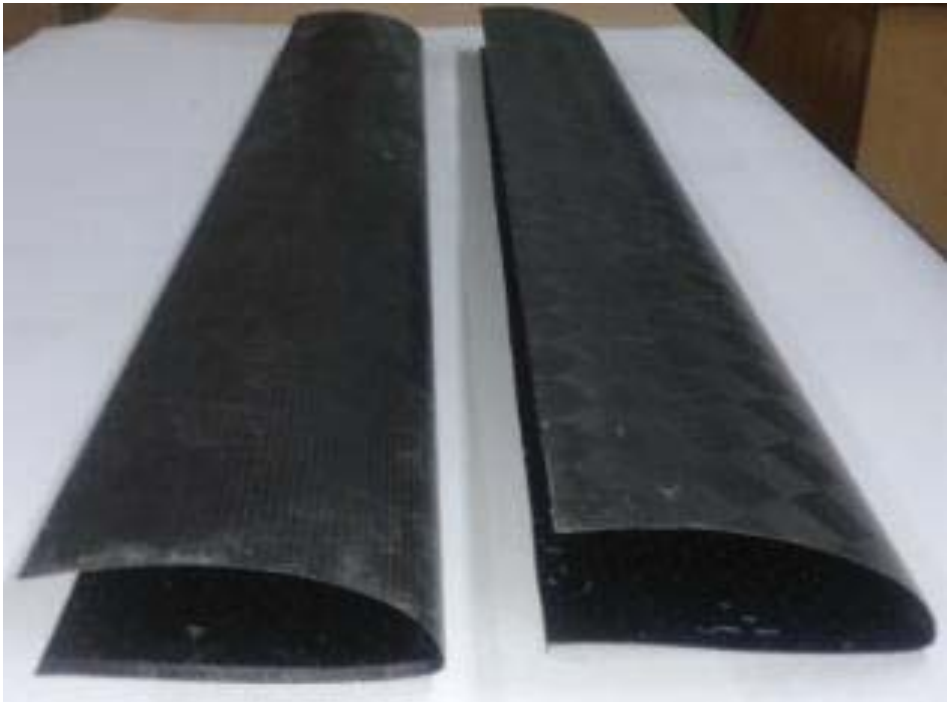


Fig. 9. Developed wing structure for ACC 2015 competition

ACKNOWLEDGEMENT

This paper was performed within a framework of project "Future Generation 2014" supported by the Polish Ministry of Science and Higher Education.

REFERENCES

- [1] CHUNG D.: *Composite Materials – Science and Application*. Springer 2010.
- [2] ŁUKASZEWICZ A.: *SolidWorks based CAX Education Directed for Industrial Practice*. SolidWorks World 2010, Anaheim, USA, p. 23.
- [3] DANKWORT C., WEIDLICH R., GUENTHER B., BLAUROCK J. E.: *Engineers' CAX education-it's not only CAD*. Computer-Aided Design, Vol. 36, 2004, pp. 1439–1450.
- [4] PLANCHARD C., PLANCHARD M.: *SolidWorks 2011 Tutorial*. Schroff Development Corporation, Mission KS, 2011.
- [5] LOMBARD M.: *SolidWorks 2010 Bible*, Wiley Publishing, Indianapolis, USA, 2010.
- [6] GRODZKI W., ŁUKASZEWICZ A.: *Design and manufacture of unmanned aerial vehicles (UAV) wing structure using composite materials*. Materialwissenschaft und Werkstofftechnik, Vol. 46, No. 3, 2015, pp. 269–278.
- [7] KARAKAS H., KOYUNCU E., INALHAN G.: *ITU Tailless UAV Design*. Journal of Intelligent and Robotic Systems, 2013, Vol. 69, p. 131.
- [8] GRODZKI W., ŁUKASZEWICZ A.: *CAX tools in the design process of an unmanned flying device. Part I Aerodynamic analysis*. Mechanik, Vol. 12, 2013, pp. 95–102.

*Medjon HYSENAJ**

INTERACTIVE METHODOLOGY FOR SPATIAL REPLICATION AND RESEARCH

Abstract

The paper outlines geographic search technology as the future of research methodology. Due to the importance of geographical environment in the routine issues, the close connection between spatial extension and most of the phenomena running over the world, the geographic query process turns to be key factor in overcoming hazard decision-making and evaluation. We describe a two step process where derived output from spatio-textual research could be largely disseminated by spatial replication techniques. Spatial research differs in that it requires a combination of text and geographical data. The mathematical formula to calculate distances derives from the assertion that the Earth is nearly spherical (ignoring ellipsoidal effects) so great-circle distance formulas give the distance between points on the surface of the Earth correct to within 0.5% or so. Based on the search range and limit, we introduce the Cosine algorithm who offers better speed performance and data retrieval accuracy. Retrieved data exploit spatial replication technique potentiality. When it comes to spatial data we deal extensive amount of information. Web Gis Albania is an informative platform initially structured for academics purposes further enriched with modules which support the spatial data circulation.

1. INTRODUCTION

The World Wide Web nowadays has turned into an endless research environment where everybody is trying to retrieve the largest amount of data within the minimum efforts. It becomes necessary to structure the appropriate platform to handle the increasing demands of the web users.

* University of Shkoder, Department of Geography, Albania, mhysenaj@unishk.edu.al

Geographic web search engines allow users to constrain and order search results in an intuitive manner by focusing a query on a particular geographic region [1]. Geographic research brings the proper information to the proper owner, avoiding abundant data out of user's interest. Content that is tailored to a region is potentially more useful to users from that region [2].

The vast amount of geographic information concerning the real world contains rough uncategorized data. Through spatio-textual research (figure 1) we aim to create a categorized repository system object to massive dissemination. The goal is to reach a unified level of geographical information acting as a standard for most users.

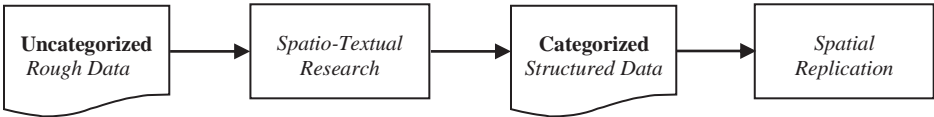


Fig. 1. Spatial-cycle data circulation [source: own study]

The terms geographic queries and spatial queries imply querying a spatially indexed database based on relationships between particular items in that database within a particular coordinate system (or compatible coordinate systems) [3]. The Web GIS Albania application (Figure 2) combines spatial with textual queries.

We exploit mathematical algorithms for spherical search coordinates mainly represented by the law of Cosine with textual queries based on the entity attributes from who derive the platform geo-database. A spatial database is a database system that offers spatial data types in its data model and query language and supports spatial data types in its implementation, providing at least spatial indexing and spatial join methods [4]. Geographic data is often subject of local or global target for multiple profile entities such as business and marketing strategy, hazard management, logistic assignment, etc.

As for large part of web research has strict implicit or explicit connection to geographic location we determine that these issues turn to be geo-sensitive. Hence the use of spatial database as the backbone repository instrument for the application could turn into a logic solution for the increasing number of geo-requests. A spatial database is considered complete from that it accomplishes thoroughly every complex request. Hence it is important enriching the geo-database with current and updated information.

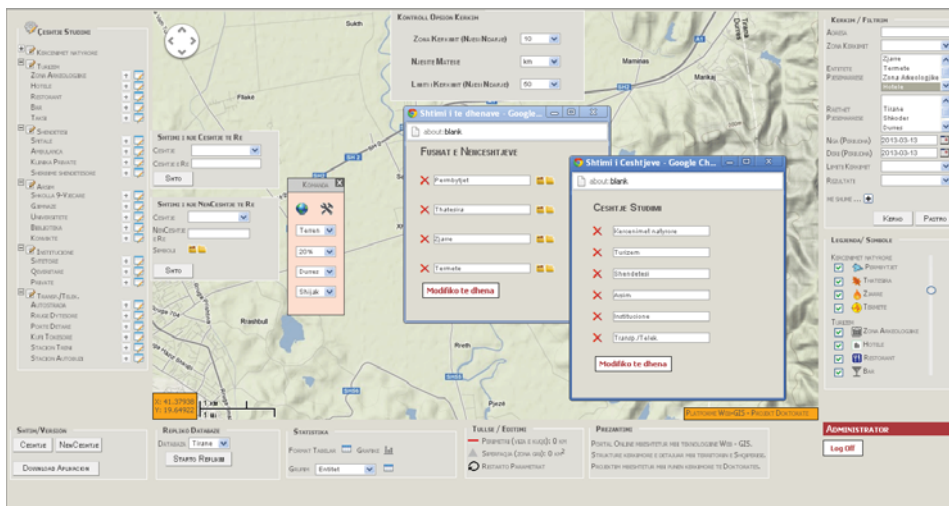


Fig. 2. WGA application, Back-office [source: own study]

2. RESEARCH METHODOLOGY

The application research process allows the user to deal with multi functionalities, covering large part of user request, without the need to perform distinct and consecutive actions. Let's analyze two distinct cases.

Case one: the request consists to retrieve a list of bars, hotels and restaurants within e distance of 10 km from a specific hotel? For management purpose the investor need to retrieve a list of these three business entities that could turn useful or maybe potential competitors fur his future investment. He isn't interested in other entities so he doesn't expect the list to contain other useless data. By so the search engine offers the exact output he needs with minimum efforts. From figure two we notice there are six entities that fulfill both requirements distance and type. Further the user needs a report containing only the bars? Although the distance remains the same, the modified query will turn already three from previous points.

Case 2: During the period 2010-2011, Shkodër district suffered from large scale flood. The population was affected from both social and economic aspect. By that time, responsible authorities were charged for uncoordinated action and slow intervention process. Basically the created situation was the product of a missing platform, able to lead to well organized instructions. The elaboration of an evacuation plan consists in finding the optimal routes (distance and security), accommodation entities (shelter for affected population) from flooded areas, etc. The current application would have solved many uncertainties in short time.

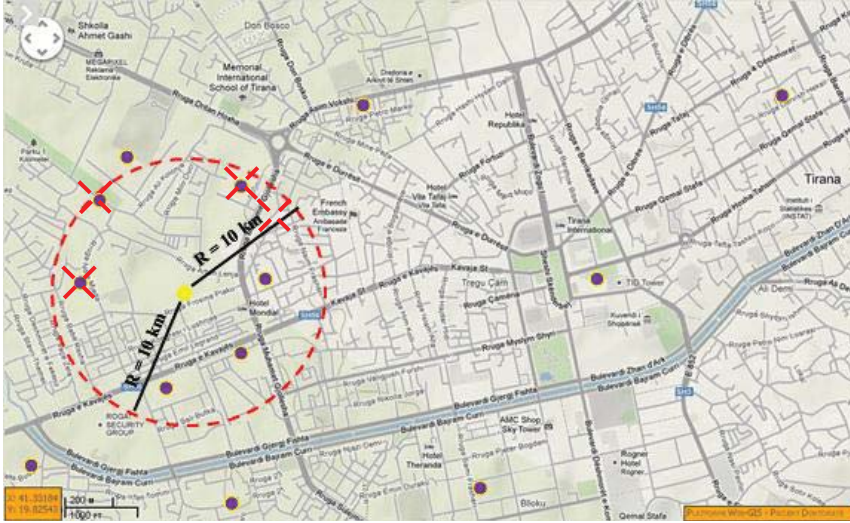


Fig. 3. Research schema

Three years ago after the first online appearance, the application handled the research process exploiting the Haversine formula (1). Principally the Haversine formula is far superior from other formulas for small distances at the range of a couple of meters. However the applications goal is to perform researches far beyond this limit.

$$dLat = (lat2-lat1) \quad (1)$$

$$dLon=(lon2-lon1)$$

$$a=\sin(dLat/2)*\sin(dLat/2)+\cos(lat1)*\cos(lat2)*\sin(dLon/2)*\sin(dLon/2)$$

$$distance = 6371 * 2 * \operatorname{atan2}(\operatorname{sqrt}(a), \operatorname{sqrt}(1-a))$$

Claiming larger distances at the range of kilometers, we discovered that Vincenty formula or the law of Cosines have better performance as for accuracy matter. The Law of Cosines (2) uses spherical geometry to calculate the great circle distance for two points on the globe [6]. Although the Haversine formula is more robust to floating point errors [5], in fact, JavaScript (and most modern computers & languages) use 'IEEE 754' 64-bit floating-point numbers, which provide 15 significant figures of precision, with this precision; the simple spherical law of Cosines formula gives well-conditioned results [6]. On the other side the law of Cosines performs much faster than the Vincenty formula. So taking into account speed performance and accuracy of retrieved data, we conclude that at present the law of Cosines meets better application goal.

$$d = \operatorname{acos}(\sin\phi_1 \cdot \sin \phi_2 + \cos\phi_1 \cdot \cos \phi_2 \cdot \cos \Delta\lambda) \cdot R \quad (2)$$

Below we notice the low of Cosines formula integrated into the application search code where the 's' variable according to the row of appearance stands for: latitude, longitude, latitude, search_radius and search_limit.

```
SELECT name, address, latitude, longitude, (6371 * acos(cos(radians('%s')) *
cos(radians(lat)) * cos(radians(lng) - radians('%s')) + sin( radians('%s')) *
sin( radians(lat)))) FROM geographic_table HAVING search_radius < '%s' ORDER BY
search_radius LIMIT 0, '%s'"
```

The interface (figure 4) containing the necessary tools to perform the required actions. The goal is to dismantle every issue into elementary parts so that the user can easily reach the target. The set of filters handle complex queries structured to cope various situations in a stressful and eager investigation environment, also based on the involved entities additional criteria have been added. They reflect relevant attributes derived from the responsible geo-database. Further the intention remains to keep enriching the platform with additional modules and research specifications.

The screenshot shows a web-based search interface titled "KERKIM / FILTRIM". It contains several input fields and dropdown menus for filtering search results. The interface is organized into sections with labels on the left and right sides.

- Search radius:** Includes a text input for "ADRESA" (Hotel Colosseo) and a dropdown for "ZONA KERKIMIT" (10 kilometer).
- Involved entities:** Includes a dropdown for "ENTITETE" (Rezerva Natyrore, Hotele, Shtepi Private, Restorant) and a dropdown for "PJESEMARRESE" (Shtepi Private, Restorant).
- District:** Includes a dropdown for "RRETHET" (Tirane, Shkoder, Durres) and a dropdown for "PJESEMARRESE" (Tirane, Shkoder, Durres).
- Time evaluation:** Includes text inputs for "NGA (DATE)" (2015-05-13) and "DERI (DATE)" (2015-05-22).
- Limit of research:** Includes a dropdown for "LIMITI KERKIMIT" (30 Entitete) and a dropdown for "REZULTATI" (empty).
- Additional search criteria:** Includes a button "ME SHUME ..." with a plus icon.
- Buttons:** At the bottom, there are two buttons: "KERKO" and "PASTRO".

Fig. 4. Research tool [source: own study]

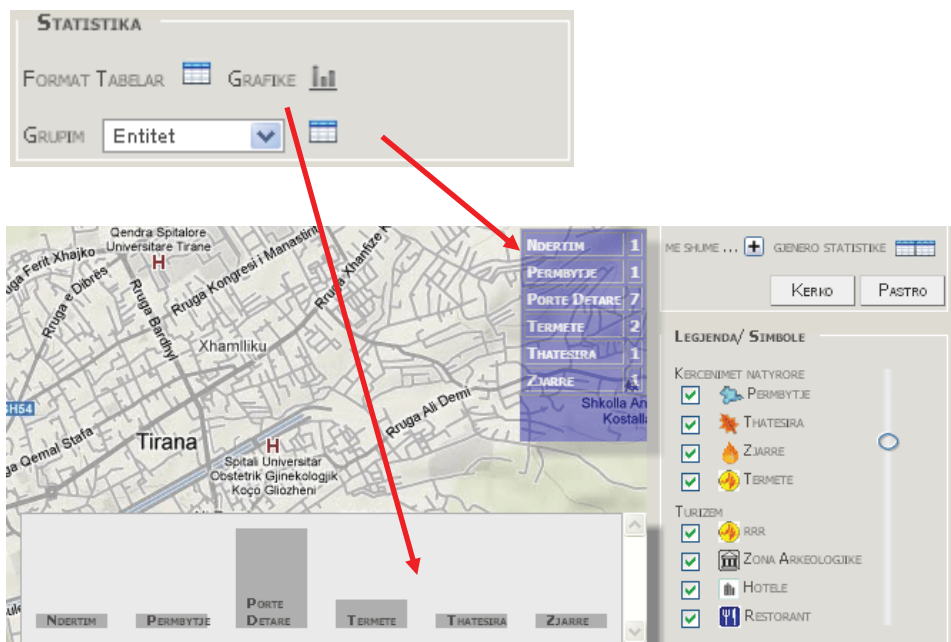


Fig. 5. Statistical tool [source: own study]

As the main goal of the researcher is to produce data in terms of reports or synopsis WGA automatically generates statistics (figure 5) based on the output. The report can be structured based on two elements entity or location. In case we scan a specific area (usually at district level) means we are interested to analyze entities for that area. In case we perform the research on multiple areas (more than one district) it is possible to group as well as by entity or location. Generated data can be analyzed on graphs or tabular formats and exported into pdf., xls. or doc. formats. The user reaches a new standard of just visualizing the current issue.

3. REPLICATION TECHNIQUE

The development of handled applications such as tablets, smart phones, laptop, for dynamic users creates optimal conditions to focus on Gis technology. Data replication stands as an important pillar for management issues for many private and government entities. Data replication is driven by the need to maintain more accurate and up-to-date spatial databases, but the same time reduce data acquisition and maintenance costs [7]. Due to the new administrative reform in the country a large number of issues need to be restructured.

The new territorial reform affects directly spatial attributes which also need to be reflected on maps. Old data must be replaced by entities. This will affect most of spatial distributed object which will need partial or full update. When it comes to geographic entities we easily presume the large quantity of data to be affected. Using traditional methods could cause process overload and inconsistent data management. Efficient data exchange is crucial to the successful future of the geospatial sector [8]. Gis web based platform for data management associated with replication techniques for data turnover and distribution produce an optimal combination to support such issue.

The spatial database contain exclusively geographical coordinates associated with relevant attributes for each object avoiding this way redundant data that could slow the application performance. It wasn't necessary to store data related to the map structure; due to internet performance has increased sharply, this allow us to exploit pre-generated maps offered by Geo suppliers which are much faster to load compared to local geo files. We exploit maps offered by Google through Google APIs by creating a second layer of markers for every accomplished research which overlays the existing map. This schema allows developing a two sided data flow between the user and the spatial database. A successful research is translated into a generated XML file containing located objects and the respective coordinates and attributes.

4. DATA FLOW SCHEMA

Nowadays creating multiple copies of information into multiple locations turns into an essential step for a healthy performance. MySQL Replication raises the bar – a whole new level in terms of performance [9]. Platforms dealing with large amount of data are closely related with such supporting techniques. Web Gis Albania primarily relies on geographical data.

Among all types of data, the necessity to keep spatial data up-to-date is imperative for their dynamic nature. We claim the term dynamic due to the rapid evolution. A restaurant may close or transfer somewhere else. Bars can change name or category. Business entities are often subject to developments. Hence information reflected geographically, associated to real world objects, evolves rapidly, as much as a week or month could be considered high value variables.

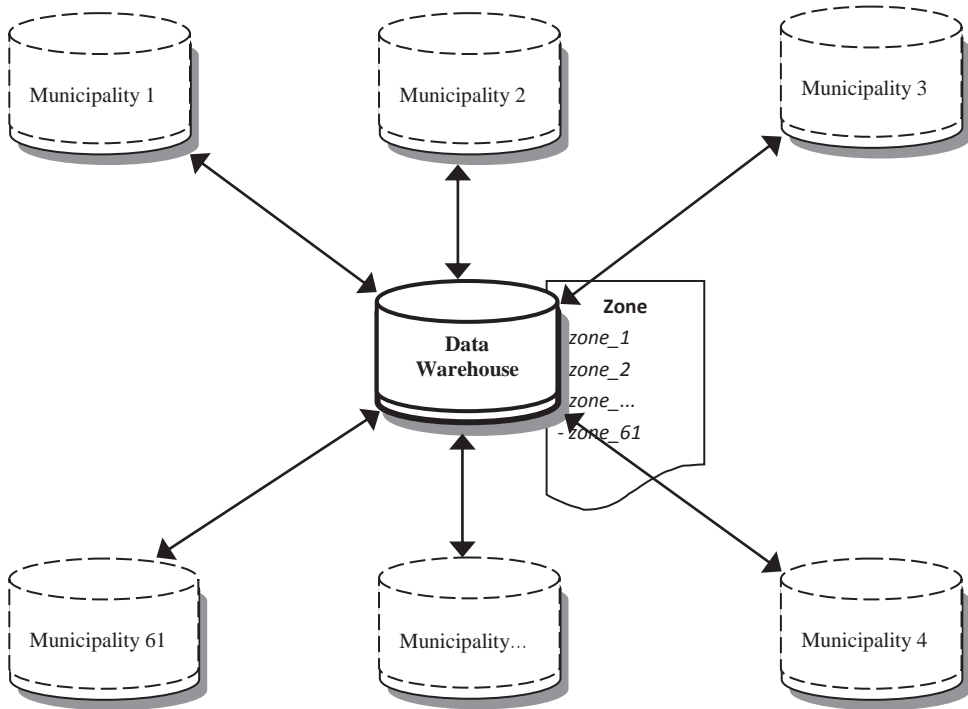


Fig. 6. Replication schema [source: own study]

The new territorial reform divides the country into 61 administrative areas called municipality. We need geographically dispersed units converging into a centralized environment. That requires 61 peripheral slave servers connected into one master server (Figure 6). Peripherals act as autonomous entity which collect, distribute and manage geographical data related to the associated area.

Each peripheral unit asks for specialists with sustainable geographical area knowledge associated with supportive skills in managing the platform system. The potentiality of WGA to generate reports has direct effect on the server performance which can lead to overload and data loss. Multiple backup and healthy recovery plans must always be on standby. This might need great efforts leading to massive consumption of physical and time sharing resources. Replication is considered a modern solution for backup and a healthy failover management.

From previous paper work [10] I described the application data flow schema (Figure 7). We use two distinct database systems. Spatial database refers to any set of data describing the semantic and spatial properties of real world phenomena (temporal properties are also possible) [11]. The so-called “spatial database” is associated the context of data which preserves. We structure spatial database responsible for all entities belonging to the analyzed area.

Data insertion occurs through a simple mechanism, who transforms objects identified in real world into coordinates which are associated with corresponding attributes. They are unique as the objects reflecting the outside reality. These coordinates constitute the groundwork from which we manage to display symbols on maps and further structure reports based on the output.

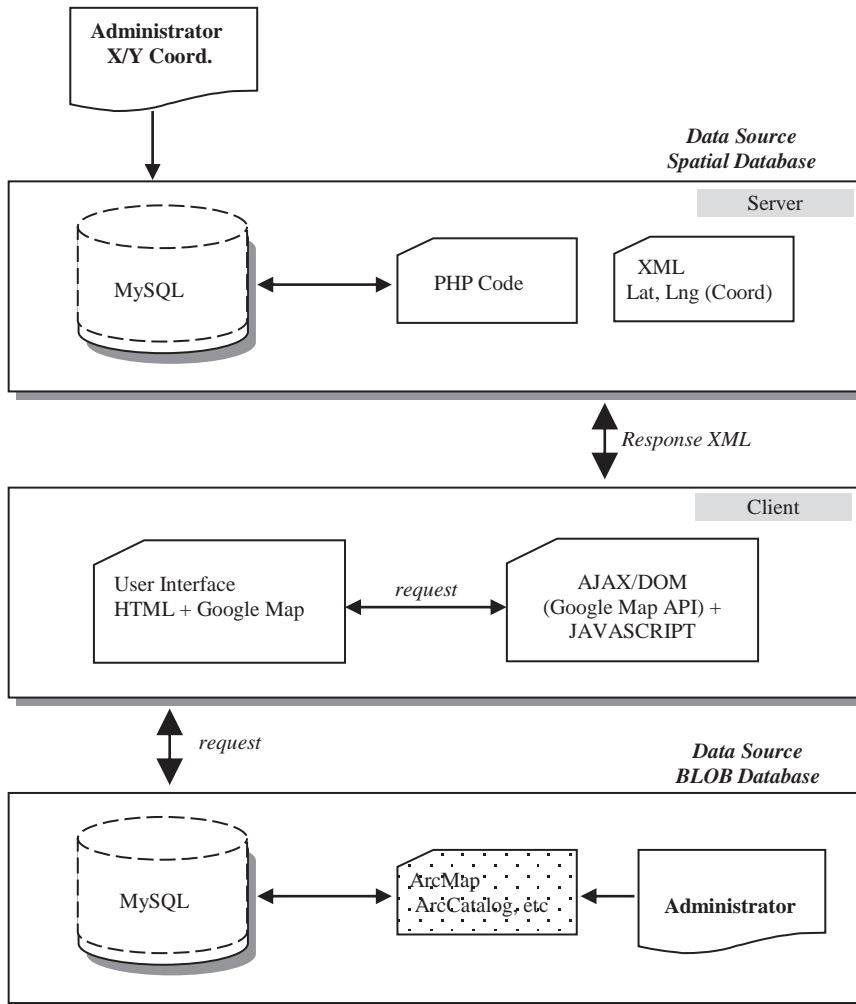


Fig. 7. Data flow schema, Web Gis Albania [source: own study]

We exploit a second database to store pre-processed dynamic maps with the help of ArcGIS package. After that, we store maps converted into images .gif or .jpeg type in the database as blob files. Blob is a binary large object that can hold a variable amount of data whose values are treated as binary strings. It is compatible when it comes to store image data [12].

Figure 8 present the spatial database schema which store any entity that can be spatially displayed and represent a point of interest for the population and decision-making authorities. Each primary entity is divided into categories in a one-to-many relationship reflecting specific issues. For each category in a one-to-many relationship, we structure databases reflecting relevant units storing coordinates and specific attributes for each object of that category. Most of information is stored in these tables.

The centralized spatial database claims the same table structure. Meanwhile peripheral databases preserve the geographical source of the derived data through the additional attribute *id_municipality*, essential to avoid data overlap. Through this column all data replicated into the Master database could be further exploited to produce general reports.

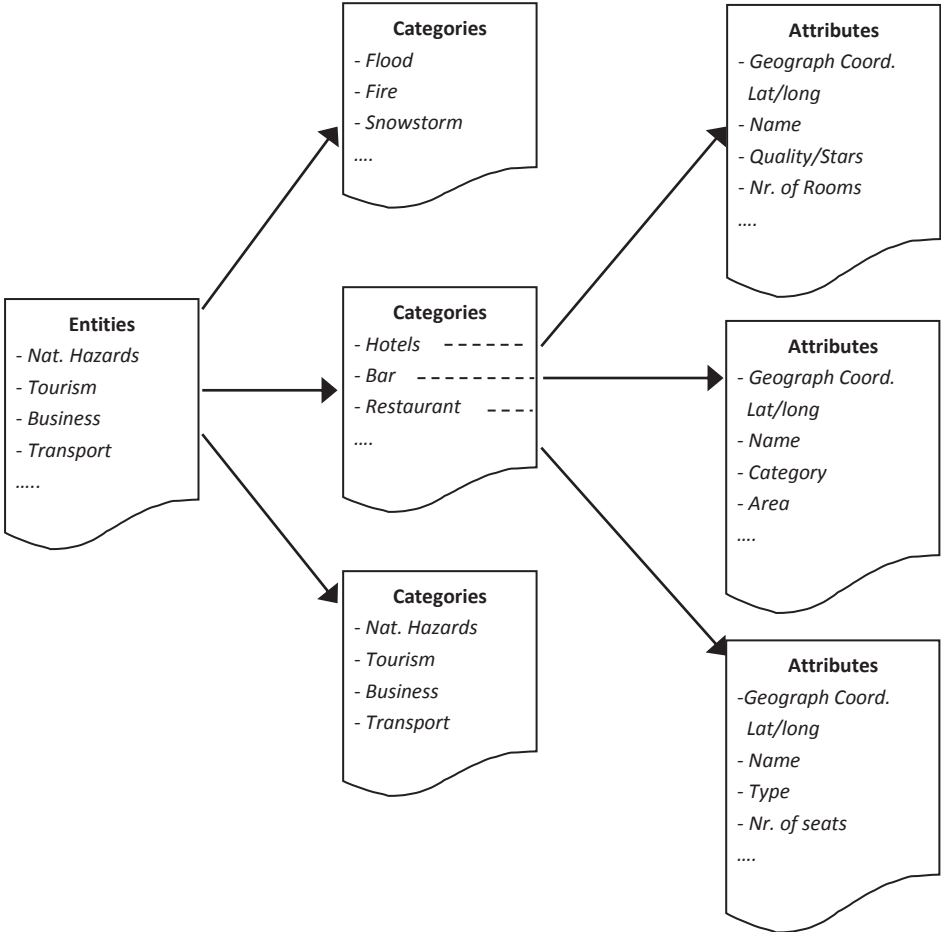


Fig. 8. Spatial database schema [source: own study]

5. CONCLUSIONS

All aspects of human activity are rooted in geographic space in some aspect [13]. As a result references to geographical locations are prevalent everywhere. The paper goal is to present a spatial cycle data circulation based on a spatio-textual research transforming rough data into structured information further disseminated by spatial replication techniques. The spherical algorithm such as the law of Cosin associated with additional text queries produced interesting results. We build a GIS based platform, initially for academics research, offering a clear perspective for a spatial search engine. On the basis of well performed query stands a fully developed spatial database. That's why the platform needs frequent "attention".

The new territorial reform is an incentive to reflect geographical evolution. Good territory administration creates accurate, accessible, interoperable, timely, secure, and complete information about land in an affordable and efficient way that promotes confidence between the public, its commercial enterprises, and government [14]. Wherever geographic evolution is present there is always necessity to reflect spatial changes. Administrative units include vast entities which by them undergo deep changes not only geographically, reflecting their spatial location evolution; also we face consistent changes in their associative attributes.

The paper addresses problems associated with spatial data replication within a data-centralized mechanism and geographically distributed environment. The scope is to settle a healthy multiple data circulation among resources whose location are spatially extended. We concentrate on object location and their geographic coordinates followed by their associative attributes. The maps act as linen where results from the spatio-textual research overlay on it. The data flow allows us to produce in short time complex data analyze and explicit reports based on decision-makers requirements.

The replication technique will allow us to act as a connection bridge for all units into a centralized repository from which initiate most of the consentaneity for the geographic data flow. Through spatial replication we aim to reduce time extension of processes, increase availability and failure management, create a reliable standard from which people can easily retrieve informative data. WGA act as a source of communication also as a tool for most academics projects. The goal is to prompt Gis supporters to rely on spatial technology and to further enrich their skills on spatial technology. Location information integrated into our digital information environment adds a previously missing conceptual dimension to our formulation of problems and questions [15]. The spatial component will turn into an underlying stone for many interactive processes such as urban planning, environmental issues, research methodology, business operations, etc. It is necessary to enlarge and upgrade the application. In the future perspective further enrichment of the spatial database will turn a key issue.

REFERENCES

- [1] YEN-YU CH., TORSTEN S., ALEXANDER M.: *Efficient Query Processing in Geographic Web Search Engines*, 2006.
- [2] KEVIN S.: *Geospatial Mapping and Navigation of the Web*. IBM Almaden Research Center, U.S.A, 2001.
- [3] RAY L.: *Geographic Information Retrieval and Spatial Browsing*. School of Library and Information Studies, University of California, Berkeley, 2000.
- [4] RALF G.: *An Introduction to Spatial Database Systems*. Praktische Informatik IV, FernUniversität Hagen, D-58084 Hagen, Germany, 1994.
- [5] Formulas to Calculate Geo Proximity, www.stackoverflow.com.
- [6] CHRIS V.: *Calculate distance, bearing and more between Latitude/Longitude points*. www.movable-type.co.uk/scripts/latlong.html
- [7] MARK S., Katherine I.: *Data Replication and Data Sharing – Integrating Heterogeneous Spatial Databases*.
- [8] SERRYN E., FRANCISCO E., IAN W.: *Hierarchical spatial reasoning theory and gis technology applied to the automated delineation of administrative boundaries*. 2000.
- [9] LUIS S., LARS T.: *MySQL Replication High Performance: Multi-Threaded Slaves and Group Commit*, 2012.
- [10] MEDJON H.: *Web GIS Albania Platform, an Informative Technology for the Albanian Territory*, Informatica 2012.
- [11] BEDARD Y.: *Principles of spatial database analysis and design*, 2005.
- [12] ORACLE, MySQL reference manual, 2015.
- [13] CHRISTOPHER B., ALIA I., DAVID F., GAIHUA F., SUBODH V.: *The SPIRIT Spatial Search Engine: Architecture, Ontologies and Spatial Indexing*. School of Computing, Cardiff University, Cardiff, Wales, CF24 3XF, UK, 2004.
- [14] ROBIN M., VICTORIA S.: *Ict for land administration and management*. 2011.
- [15] The Spatial Web, An Open GIS Consortium (OGC) White Paper, www.opengis.com, 2004.

Iryna KARAYUZ, Petro BIDYUK***

FORECASTING GDP GROWTH RATE IN UKRAINE WITH ALTERNATIVE MODELS

Abstract

The problem of constructing mathematical model for short-term forecasting of GDP is considered. First, extended autoregression is constructed that takes two additional independent variables into consideration. The model resulted provides a possibility for generating short-term forecasts of GDP though not of high quality. Another model was constructed in the form of a Bayesian network. The model turned out to be better than the multiple regression, it provides quite good estimates for probabilities of GDP growth direction.

1. INTRODUCTION

Gross domestic product (GDP) is an integrated macroeconomic process that together with level of inflation characterizes to quite acceptable extent current state of macroeconomics. There exists a substantial interest to modeling and forecasting GDP with hiring various model types [1–3]. In [1] the issues are analyzed of pooling models for a given set of individual N units observed over T periods of time. It is shown that the forecasting results received by the authors exhibit high quality and outperform ordinary least squares based forecasts. The generalized factor model with infinite dynamics and non-orthogonal idiosyncratic components is considered in [2]. The authors constructed the coincident index for European Union. The dynamic factor model for forecasting the euro area GDP from monthly indicators is developed in [3]. It was found here that surveys and financial data contain important information for the GDP forecasts beyond the monthly real activity measures.

* Postgraduate student, National Technical University of Ukraine “Kyiv Polytechnic Institute”, Institute for Applied System Analysis, 03056, building № 35, 37 Prosp. Peremohy, NTUU “KPI”, Kiev, Ukraine, +38 (050) 145-72-54, afanasievaiv@ya.ru

** Professor, Dr. of Eng. Sci., National Technical University of Ukraine “Kyiv Polytechnic Institute”, Institute for Applied System Analysis, 03056, building № 35, 37 Prosp. Peremohy, NTUU “KPI”, Kiev, Ukraine, +38 (097) 213-47-08, pbidyuke@gmail.com

As an indicator of inflation process very often is used well known consumer price index (CPI) [4]. To determine correct current estimate of macroeconomic state and to evaluate short-term forecast estimates for economic process of interest it is necessary to construct mathematical models using historical statistical data. At the same time the use of well substantiated mathematical models and forecasts that are based on them does not exclude the possibility of quality expert estimates usage. These estimates could be useful in modeling so that to determine initial conditions or prior probabilities, to form restrictions on key variables of interest, to select appropriate modeling techniques, to estimate elements of model structure, hidden variables etc [5].

The forecasting models for financial and economic processes can be of different complexity level. A wide class of such models may include the following types: multiple regression, autoregression with moving average (ARMA), autoregression with integrated moving average (ARIMA), nonlinear and polynomial regressions as well as numerous modifications and combinations of the structures mentioned. Also highly popular today are the models constructed with the use of artificial intelligence techniques such as neural networks, fuzzy logic, neuro-fuzzy models, a wide class of Bayesian type models (Bayesian regression and Bayesian networks), and the models built with application of support vector machine methodology [6, 7].

2. PROBLEM STATEMENT

The main task of the study is to construct mathematical model describing growth rate for the Ukrainian GDP. Dynamic Bayesian network should be constructed reflecting existing cause-and-effect links between selected key variables and generating short term probabilistic forecasts. Finally a comparison of forecasts estimates should be performed computed with different model types.

3. ANALYSIS OF CURRENT SITUATION IN UKRAINE

Using the economic production methodology GDP could be found as a sum of gross value added (GVA) for all kinds of economic activities plus net tax [8]. The model should take into consideration growth rates for economy branches, consumer price index (CPI), production price index (PPI), GVA structure changes etc.

Actual income of population has decreased in 2014 by 8.4%; population's savings decreased by 26.5%; CPI increased by 24.9%; PPI for industrial production increased by 17.1% [9].

The dynamics of structure of goods and services production for separate economic activities is changing as shown in Fig. 1 [9]. The GVA structure is influenced by the following factors:

- changes of law (for example, rate of rent for natural gas mining for joint ventures is 70% from the cost of final product [9] what negatively influences attractiveness of such activities in 2015);
- local war (LW) in the south-east of Ukraine (Lugansk and Donetsk areas) has highly negatively influenced the volume of industrial production of these areas and Ukraine as a whole due to existing links between enterprises;
- temporary occupation of Autonomous Crimean Republic, Sebastopol area.

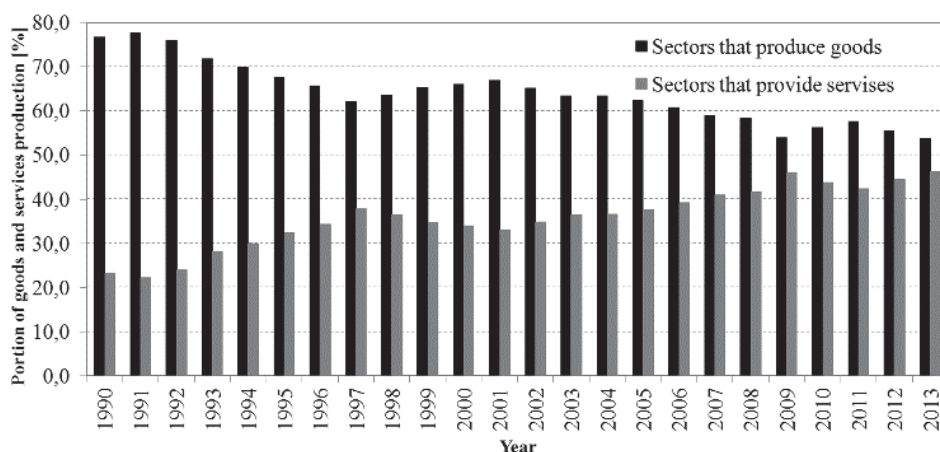


Fig. 1. Dynamics of the structure of goods and services production [source: own study]

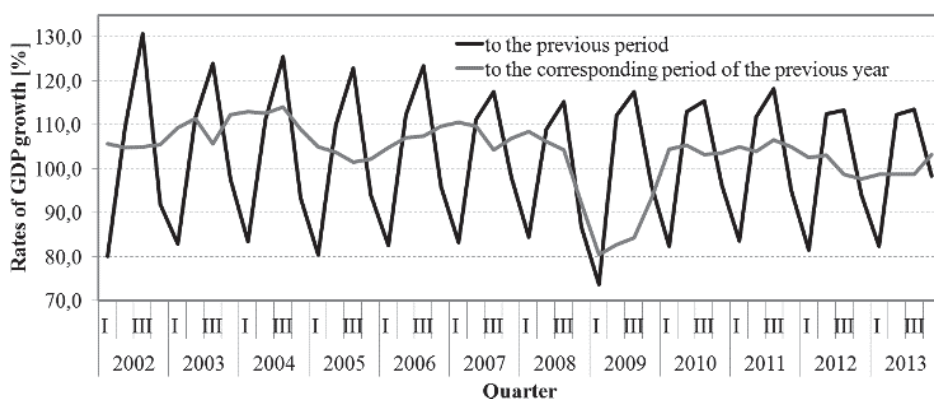


Fig. 2. Rates of GDP growth, quarterly data [source: own study]

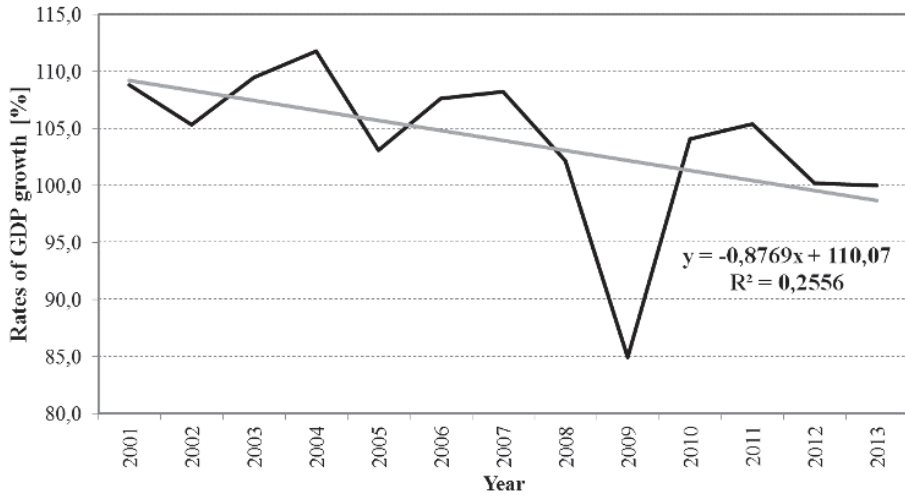


Fig. 3. Rates of GDP growth GDP, annual data [source: own study]

Figs. 2 and 3 rates of growth for GDP with respect to the same period of previous year. Each factor is informative for modeling and to reach quality forecasting it is necessary to take into consideration all the time series mentioned.

4. REGRESSION MODEL

First, consider the possibility for constructing model in the form of extended autoregression (ARX):

$$y(k) = a_0 + \sum_{i=1}^p a_i y(k-i) + \sum_{j=1}^q b_j x(k-j) + \varepsilon(k),$$

where: p – autoregression order,
 q – a number of regressors,
 $y(k)$ – the main dependent variable at discrete moment of time k ,
 $x(k)$ – regressor,
 a_i – i -th autoregression part parameter,
 b_j – j -th parameter of multiple regression part of the model,
 $\varepsilon(k)$ – random process that is formed by nonmeasurable stochastic external disturbances, measurement errors, parameter computing errors, and model structure inadequacy.

Such models provide a possibility for constructing forecasting functions that enable computing of forecasts for a necessary number of steps using conditional mathematical expectation operator.

According to the correlation analysis performed the model should include CPI and PPI with lags 3 and 2, respectively:

$$GDP(k) = 75,8 + 0,94 \cdot GDP(k-4) - 0,21 \cdot GDP_4(k-4) - 0,15 \cdot CPI(k-3) - 0,31 \cdot PPI(k-2).$$

where: GDP is relative GDP growth to the previous time period [%],
 GDP_4 is relative GDP growth to the same time period of previous year [%],
 CPI is CPI growth with respect to the previous time period [%],
 PPI is PPI growth with respect to the previous time period [%].

Adequacy of this model is characterised by the following statistics: $R^2 = 0,95$; $DW = 1,55$; $SSE = 429,1$. The parameters estimates correspond to the realities of GDP changes in time: when CPI and PPI are growing GDP growth rate is decreasing.

The forecasts quality is characterized by the following statistics: $MSE = 389,7$; $MAPE = 16,8\%$; $U = 0,13$ (Theil coefficient).

Tab. 1. Comparison of actual GDP (SSSU) and its forecast for 2014

Time period	GDP [%]		Forecast error [%]
	SSSU	Model	
QI 2014	76,4	83,2	8,91
QII 2014	108,4	111,1	2,52
QIII 2014	112,0	110,8	-1,03
QIV 2014	88,8	91,8	3,34
2014 year	93,2	100,2	7,55

The statistical data in Table 1 were taken from the State Statistical Service of Ukraine (SSSU) as of March 20, 2015 [9].

As far as mean absolute percentage error is about 17% for the regression ARX model considered above it is necessary to try another approach to modeling GDP that would be ideologically different. Besides, this model cannot describe quantitatively and qualitatively influence of the war in the south-east of Ukraine. Also one of important influence factors to development of GDP is support of the dollar/hryvna exchange rate practically at fixed level) from May 2012 till February 2014. Define this event as “Pseudo Stability).

5. BAYESIAN NETWORK CONSTRUCTION

Bayesian network (BN) is a directed acyclic graph vertices of which are model variables, and its arcs show existing cause-and-effect links between the variables. To construct a model in the form of a Bayesian network introduce the following variables: Local War (LW) is a variable that takes value of “1” when there is a war on Ukrainian territory (southeast); Pseudo Stability (PS) takes a value of “1” if exchange rate for hryvna (Ukrainian currency) is fixed; Investment Climate (IC) takes value of “1” if the rate of investments growth into Ukraine is positive; CPI(-3) and PPI(-3) are equal “1” if their growth rate exceeds 15%; GDP is equal “1” if its growth rate is positive.

The first BN was built by hiring expert estimates and with the use of model quality criterion called minimum description length (MDL), it is shown n Fig. 4. The models given in Fig. 5 were constructed by heuristic algorithm [10] with the use of computed values for mutual information in the following way: a) without restrictions; b) with restriction that variable 6.GDP can be only subordinated (child), and the variables “1. LW”, “2. PS” can only be of parents type; c) with restrictions of type b) and without free vertices.

The net given in Fig. 5a cannot be used practically because GDP growth rate cannot influence growth rate of PPI that was three quarters (9 months) before.

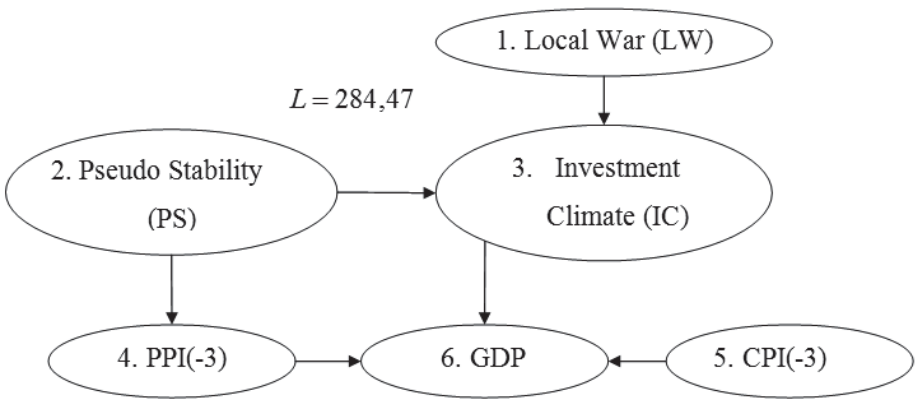


Fig. 4. Network for GDP of Ukraine built by expert [source: own study]

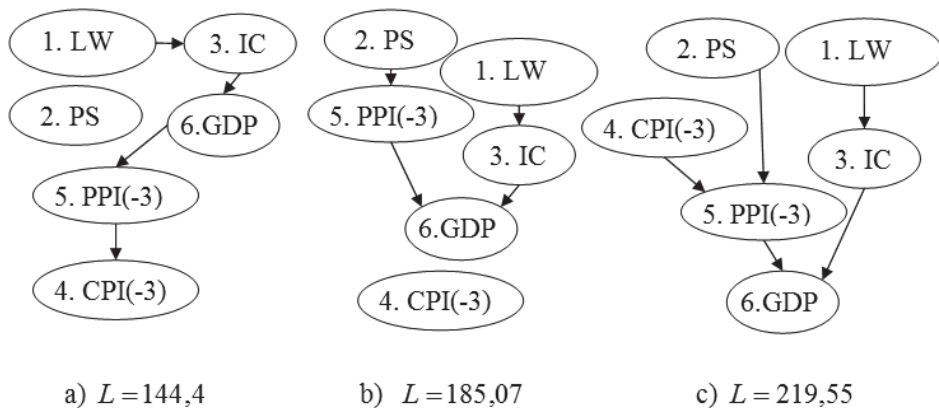


Fig. 5. Networks for GDP of Ukraine built by heuristic algorithm: a) without restrictions; b) with restriction that variable 6.GDP can be only subordinated (child), and the variables “1. LW”, “2. PS” can only be of parents type; c) with restrictions of type b) and without free vertices [source: own study]

P(PS)		P(LW)	
PS=0	PS=1	LW=0	LW=1
0,85	0,15	0,25	0,75

	P(PPI(-3)/PS)			P(IC/LW)	
	PPI(-3)=0	PPI(-3)=1		IC=0	IC=1
PS=0	0,49	0,51	LW=0	0,44	0,56
PS=1	1	0	LW=1	1	0

		P(GDP/PPI(-3),IC)	
		GDP=0	GDP=1
PPI(-3)=0	IC=0	0,67	0,33
PPI(-3)=0	IC=1	0,07	0,93
PPI(-3)=1	IC=0	0,91	0,09
PPI(-3)=1	IC=1	0,70	0,30

Fig. 6. Conditional probability tables for BN in Fig. 5b [source: own study]

Expression for computing joint probability function for BN in Fig. 6:

$$P(PS, LW, PPI(-3), IC, GDP) = \sum \{P(GDP|PPI(-3), IC) \cdot \sum \{P(PPI(-3)|PS) \cdot P(PS)\} \sum \{P(IC|LW) \cdot P(LW)\}\}$$

P(PS)		P(CPI(-3))		P(LW)	
PS=0	PS=1	CPI(-3)=0	CPI(-3)=1	LW=0	LW=1
0,85	0,15	0,875	0,125	0,25	0,75

		P(PPI(-3)/PS,CPI(-3))		P(IC/LW)	
		PPI(-3)=0	PPI(-3)=1	IC=0	IC=1
PS=0	CPI(-3)=0	0,54	0,46	LW=0	0,44 0,56
PS=0	CPI(-3)=1	0,17	0,83	LW=1	1 0
PS=1	CPI(-3)=0	1,00	0,00		
PS=1	CPI(-3)=1	0,50	0,50		

		P(GDP/PPI(-3),IC)	
		GDP=0	GDP=1
PPI(-3)=0	IC=0	0,67	0,33
PPI(-3)=0	IC=1	0,07	0,93
PPI(-3)=1	IC=0	0,91	0,09
PPI(-3)=1	IC=1	0,70	0,30

Fig. 7. Conditional probability tables for BN in Fig. 5c [source: own study]

Expression for computing joint probability function for BN in Fig. 7:

$$P(PS, CPI(-3), LW, PPI(-3), IC, GDP) = \sum \{P(GDP|PPI(-3), IC) \cdot \sum \{P(PPI(-3)|PS, CPI(-3)) \cdot P(PS) \cdot P(CPI(-3))\} \sum \{P(IC|LW) \cdot P(LW)\}\}$$

P(PS)		P(LW)	
PS=0	PS=1	LW=0	LW=1
0,85	0,15	0,25	0,75

		P(PPI(-3)/PS)		P(IC/LW,PS)		P(CPI(-3))	
		PPI(-3)=0	PPI(-3)=1	IC=0	IC=1	CPI(-3)=0	CPI(-3)=1
PS=0	0,49	0,51	LW=0	PS=0	0,39 0,61	0,875	0,125
PS=1	1	0	LW=0	PS=1	0,71 0,29		
			LW=1	PS=0	1,00 0,00		
			LW=1	PS=1	1,00 0,00		

		P(GDP/PPI(-3),IC,CPI(-3))	
		GDP=0	GDP=1
PPI(-3)=0	IC=0	0,67	0,33
PPI(-3)=0	IC=0	0,50	0,50
PPI(-3)=0	IC=1	0,07	0,93
PPI(-3)=0	IC=1	0,00	1,00
PPI(-3)=1	IC=0	0,83	0,17
PPI(-3)=1	IC=0	1,00	0,00
PPI(-3)=1	IC=1	0,70	0,30
PPI(-3)=1	IC=1	0,50	0,50

Fig. 8. BN with nodes as conditional probability tables for BN on fig. 4 [source: own study]

Expression for computing joint probability function for BN in Fig. 8:

$$\begin{aligned}
& P(PS, CPI(-3), LW, PPI(-3), IC, GDP) = \\
& = \sum \{P(GDP|PPI(-3), IC, CPI(-3)) \cdot \sum \{P(IC|LW, PS) \cdot P(LW) \cdot P(PS)\} \cdot \\
& \cdot \sum \{P(PPI(-3)|PS) \cdot P(PS)\} \cdot P(CPI(-3))\}
\end{aligned}$$

Using the joint probability function and Bayes' theorem we got the following CPTs:

Tab. 2. Conditional probabilities for constructing BNs

	BN of Fig. 6b		BN of Fig. 6c		BN of Fig. 5	
	P(GDP/LW)		P(GDP/LW)		P(GDP/LW)	
	GDP=0	GDP=1	GDP=0	GDP=1	GDP=0	GDP=1
LW=0	0,534	0,466	0,535	0,465	0,510	0,490
LW=1	0,833	0,167	0,833	0,167	0,737	0,263

Thus, the least possible probabilistic estimate for the negative direction of Ukrainian GDP growth in 2014 is 0.74 by condition that takes place local war. Respectively the most possible probabilistic estimate for the positive direction of Ukrainian GDP growth in 2014 is 0.26.

6. CONCLUSIONS

It was shown that application of regression models for describing the processes of economy in transition not always finalizes with positive results. This can be explained by numerous out-of-market events (factors) that influence development of the economy. The short-term forecasting results obtained in this case are not of very high quality though acceptable.

On the other side probabilistic models such as Bayesian networks provide a possibility for obtaining well substantiated and quite good probabilistic estimates for the direction of GDP growth in Ukraine. A substantial advantage of the simple heuristic method used for constructing BN is in its transparency and a small number of operations. Disadvantage of the heuristic approach is that after every increase in a number of variables it is necessary to reconstruct the model graph (model structure). The minimum description length based algorithm operates according to the principle of entropy minimization (the higher is dependence between the variables the better is the case from the modeling point of view). Although such approach may result in some paradoxes regarding final model quality.

In the future we will come to the new modeling problem for quantitative estimation of Ukrainian GDP growth rate: the regression model constructed in the paper should be modified with the results of structural analysis for GVA, and the interest rate of National Bank should also be taken into consideration.

REFERENCES

- [1] HOOGSTRATE A., PALM F., PFANN G.: *Pooling in Dynamic Panel-Data Models: An Application to Forecasting GDP Growth Rates*. Journal of Business & Economic Statistics, 2000, Vol. 18, No. 3, pp. 274–283.
- [2] FORNI M., HALLIN M., LIPPI M., REICHLIN L.: *The Generalized Dynamic-Factor Model: Identification and Estimation*. The Review of Economics and Statistics, 2000, Vol. 82, No. 4, pp. 540–554.
- [3] BANBURA M., RÜNSTLER G.: *A Look into the Factor Model Black Box: Publication Lags and the Role of Hard and Soft Data in Forecasting GDP*. International Journal of Forecasting, 2011, Vol. 27, No. 2, pp. 333–346.
- [4] BRANSON W.: *Macroeconomic theory and policy*. New York: Harper & Row Publishers, 1989, p. 656.
- [5] DELURGIO S.A.: *Forecasting principles and applications*. McGraw-Hill, Boston 1998, p. 802.
- [6] RUTKOWSKI L.: *Metody i techniki sztucznej inteligencji*. Wydawnictwo Naukowe PWN, Warszawa 2005, p. 520.
- [7] POLE A., WEST M., HARRISON J.: *Applied Bayesian forecasting and time series analysis*. Chapman & Hall/CRC, New York 1994, p. 409.
- [8] *Macroeconomics* / Ed. by V.D. Basilevych, Znannya, Kyiv 2007, p. 703.
- [9] The State Statistical Service of Ukraine: <http://www.ukrstat.gov.ua>
- [10] ZGUROWSKY M.Z., BIDYUK P.I., TEREITYEV O.M., PROSYANKINA T.I.: *Bayesian networks and decision trees*. Edelveis, Kyiv 2015, p. 360.

INSTRUCTIONS FOR AUTHORS

1. The journal „Applied Computer Science” Publisher original papers written in English.
2. The published papers paper must concern problems of computer applications in production engineering, economy and management or technology.
3. The paper should be prepared using *Paper template* and be not shorter than 8 and not longer than 14 pages.
4. Units of measurement should be presented simply and concisely using System International (SI) units.
5. Manuscripts should be sent together with the filled in *Submission form* by one of the authors of the manuscript to the Editorial Board's e-mail address: acs@pollub.pl.
6. Papers must be submitted on the understanding that they have not been published elsewhere and are not currently under consideration by another journal published by any other publisher.
7. The submitting author is responsible for ensuring that the article's publication has been approved by all the others coauthors.
8. The Editorial Board sends the acknowledgment of the submissions. Only an acknowledgment from the Editorial Board officially establishes the date of receipt.
9. The Editor Board inform authors about the progress of assessment and publication processes. The publisher will do everything possible to ensure prompt publication.
10. After receiving positive reviews and after acceptance of the paper for publication, the text must be submitted in a Microsoft Word document format. Drawings and photos should be additionally submitted in the form of graphical files in the *.tif, *.jpg, *.gif or *.cdr formats. Bitmap images should be of 300 dpi resolution at least unless the resolution is intentionally set to a lower level for scientific reasons.
11. The Editorial Office reserves the right to abridge and adjust the manuscripts.
12. Authorship of the paper should be limited to those who have contributed substantially to the work.

13. Guest authorship and ghostwriting is unacceptable and will be retraced if any detected cases will expose, with the academics being reported to their institutions.
14. The contents of the articles are ultimately the responsibility of the contributor. Accordingly, the Publisher and the Editorial Board accept no liability whatsoever for the consequences of any inaccurate or misleading data, opinion or statement.

ASSESSMENT PROCESS

1. All submitted papers are subject to peer review and are expected to meet standards of academic excellence.
2. Manuscript are accepted for publication by:
 - The Subject Redactor,
 - Two independent reviewers from outside the institution of the author/authors, appointed by the Scientific Board,
 - The Language Redactor.
3. All submitted papers firstly are sent to the proper Subject Redactor (according to the author's indication of the paper's scientific area) who decides if the manuscript is relevant to the subject of Applied Computer Science. Negative decision of the Subject Redactor is tantamount to rejection of the manuscript and ends the assessment procedure.
4. Reviewers are selected in accordance with the principle of avoiding conflict of interests (examples of conflicts of interest include a personal relationship between reviewer and author, occupational subordination, direct scientific cooperation).
5. Reviewers prepare the reviews using the *Review form*. The review is provided in written form and concludes with an explicit recommendation to accept or reject the paper for publication.
6. Reviewers do not know the name of paper's author/authors (reviewer receives the text of the manuscript signified by *ID-text-number* given by the Editors)
7. The names of the reviewers for the individual publications are not revealed and are anonymous to the authors.
8. If a negative review is given, the submission is rejected.
9. The reviewer can discuss the manuscript with the Statistical Redactor or/and Language Redactor.
10. On receiving the required number of reviews, the Editors pass them immediately to authors. The Editors require that authors prepare answers to reviewers' comments, which are then sent to referees.
11. When the answers to reviewers' comments are accepted by reviewers, the Editorial Board makes a decision of publishing the manuscript. Decision with the expected date of publication is sent to the submitting author.

“GHOSTWRITING” PROCEDURE

(based on guidelines of the Ministry of Science and Higher Education, Poland)

Introduction

Reliability is one of the foundations in the education. Readers should be assured that the author's results of work are transparent, fair and honest. Authors should present works in such a way whether they are the direct authors, or used in support of a natural or legal person. Therefore, the provide information about entities that contribute to the formation of publication (their substantive or financial contribution, etc.) is necessary. It is a sign of good morals and corporate social responsibility. Opposite behaviors are the phenomenon of “ghostwriting” and “guest authorship”.

“Ghostwriting” and “guest authorship”

“Ghostwriting” is the concealment participation of a person having a significant contribution to the formation of publication. If the author's contribution in the formation of the publication is negligible or even not taken place, and it is still included as an author or co-author – then we have to deal with the phenomenon of “guest authorship”.

Procedure

In order to prevent the “ghostwriting” and “guest authorship” cases authors of the publication should reveal the contribution of individual authors in the creation of the publication. They should provide information about the authorship of ideas, assumptions, methods, etc. used for the development of publication. The main responsibility for the article bears author.

Moreover the authors should present information about sources of funding of publications, contribution of research institutions, associations and other entities (“financial disclosure”).

“Ghostwriting” and “guest authorship” are forms of scientific misconduct and therefore each detected case will be exposed. The editors will notify the institutions employing authors and scientific societies. All cases of scientific misconduct (especially violations and breaches of ethics applicable in science) will be documented.

