Central European Researchers Journal

A Word of Welcome from the Editors

Dear Colleagues, Readers and Authors,

We are very glad to present next number of the "Central European Researchers Journal (CERES)". The topics of this number cover different areas and problems in engineering and in information technology in particular. Besides the thematic criterion, all accepted papers can be divided into three groups: regular papers, student's papers and methodological papers.

The significant investigations are presented and discussed in the regular papers:

- L.Ozirkovskyy, A.Mashchak, O.Shkiliuk, S.Volochiy, "The Algorithm of Automated Development of Fault Trees for Safety Exploitation Assessment of Complex Technical Systems",

- D.A.Viattchenin, "A Technique for Outlier Detection Based on Heuristic Possibilistic Clustering",

- T.O.Hovorushchenko, "Models and Methods of Evaluation of Information Sufficiency for Determining the Software Complexity and Quality Based on the Metric Analysis Results",

- V.Butenko, O.Odarushchenko, V.Kharchenko, V.Moskalets, E.Odarushchenko, O.Strjuk, "Application of Markov Modeling for Safety Assessment of Self-Diagnostic Programmable Instrumentations and Control Systems",

- E. Parso, D. Suroviak, J. Ondrejak, M. Dragula, M. Jesensky, P. Sensel, P. Sedlacek, M. Vaclavkova, "FRIMAN",

- N.F. Agamy, "Effect of Boiling and Microwave Cooking on some Antioxidant Compounds in Highly Consumed Vegetables in Egypt".

Investigations of students illustrate papers:

- P.Rusnak, "Parser of Input Data in Reliability Analysis based on Logical Differential Calculus",

- J.Rabcan, M.Zhartybayeva, "Classification by Ordered Fuzzy Decision Tree".

The methodological aspects of implemented project discussed authors of the papers:

- U.A.Beizerau, "Approaches to the Integration of the System of Education of Belarus into the World Educational Space",

- A.S.Sokolov, O.V.Kovaleva, "Application of GIS-Technologies in Ecological Education".

With best wishes

Prof. Vyacheslav Kharchenko
Prof. Elena Zaitseva

Reviewers in 2016

Androulidakis Iosif, University of Ioannina, Greece

Cariow Aleksandr, West Pomeranian University of Technology Szczecin, Poland

Cimrak Ivan, University of Zilina, Slovakia

Frenkel Ilia, Sami Shamoon College of Engineering Bialik, Israel

Kharchenko Vyacheslav, National Aerospace University KhAI, Ukraine

Koshkin Gennady, Tomsk State University, Russia

Kostolny Jozef, University of Zilina, Slovakia

Kvassay Miroslav, University of Zilina, Slovakia

Levashenko Vitaly, University of Zilina, Slovakia

Liauchuk Viktar, Gomel State University, Belarus

Lukac Martin, Nazarbayev University, Kazakhstan

Matiasko Karol, University of Zilina, Slovakia

Slavinskaya Elena, National Transport University, Ukraine

Stankevich Sergey, Scientific Centre for Aerospace Research of the Earth, National Academy of Sciences of Ukraine, Ukraine

Subbotin Sergey, Zaporozhye National Technical University, Ukraine

Tatur Michail, Belarusian State University of Informatics and Radioelectronics, Belarus

Tushnytskyy Ruslan, Lviv Polytechnic National University, Ukraine

Vaclavkova Monika, University of Zilina, Slovakia

Volochiy Bogdan, Lviv Polytechnic National University, Ukraine

Yakovyna Vitaliy, Lviv Polytechnic National University, Ukraine

Zaitseva Elena, University of Zilina, Slovakia

CONTENTS

The Algorithm of Automated Development of Fault Trees for Safety Exploitation Assessment of Complex Technical Systems

L.Ozirkovskyy, A. Mashchak, O. Shkiliuk, S. Volochiy

Abstract — this article presents an algorithm, which allows to automate the developing of fault trees for the safety exploitation assessment of complex technical systems. As result, this algorithm produces logical and graphical representations of fault tree. Obtained fault tree can be used for the exploitation safety assessment. Presented algorithm is a further step in the development of the technology of modeling the discrete-continuous stochastic systems, based on so-called structural-automaton models.

Keywords — safety, reliability, safety engineering, fault tree analysis, complex technical system, algorithms, structural-automaton model.

I. INTRODUCTION

For the effective functioning of the complex technical system (henceforth named CTS), it's extremely important to provide the required levels of reliability and safety. The failure of CTS can have devastating consequences that can lead to significant damage, including the loss of lives. Therefore, problems of safety, reliability and failure consequences assessment, during the design stage, are very important. The assessment of the reliability and exploitation safety of CTS is carried out by modeling, which is used for determining reliability and exploitation safety indicators.

There are several approaches for reliability and exploitation safety assessment, namely: simulation modeling (hereafter SM), stochastic simulation and state space modeling. Stochastic simulation has been indicated as being the most commonly used approach [1]. Important feature of stochastic simulation is that algorithms for developing models are well-formalized, so they are simple to program, and there are quite a lot of software based on that algorithms.

The most commonly used method of stochastic simulation of CTS involves the fault tree (henceforth named FT) development [2]. FT is a type of structure flowchart, which is used for graphical representation of events (determined with deductive method) that can lead to catastrophic system failure [3]. FT analysis can be used to obtain minimal cut sets (hereafter MCS). MCS is a minimal combination of events that leads to system failure. If any event is removed from the MCS, the remaining events collectively are not able to cause the system failure. MCS, obtained via FT analysis, can also include the probability of MCS (probability of all the events from the cut set occur in the same time).

II. REVIEW OF EXISTING APPROACHES OF FAULT TREE DEVELOPMENT

In modern literature, there are many works regarding FT development and analysis. Articles [2-5] describe general principles and features of FT development. Articles [6-12] present approaches of FTs development and analysis, with their further application to specific systems.

In [6] authors present an approach for dynamic FTs development and analysis. A positive feature of this approach is that it considers characteristics such as time dependencies and repeated events. After dynamic FT is developed, authors use Monte-Carlo SM method for

L. Ozirkovskyy, Lviv Polytechnic National University, Lviv, Ukraine (e-mail: lozirkovsky@lp.edu.ua)
A. Mashchak, Lviv Polytechnic National University, Lviv, Ukraine (e-mail: maschak@lp.edu.ua)
O Shkiliuk, Lviv Polytechnic National University, Lviv, Ukraine (e-mail: shkiliuk@lp.edu.ua)
S. Volochiy, Lviv Polytechnic National University, Lviv, Ukraine (e-mail: volochiy.s@gmail.com)

further analysis of the system. It should be noted, however, that the usage of SM for dynamic FT analysis significantly increases the time costs in comparison with static FT analysis.

Article [7] presents an approach for the development of the static FT with time dependencies. Approach is described with an example of development and analysis of the FT model of the telecommunication network.

Paper [8] presents a method for FTs development and analysis, based on decomposition. In this method, FT is decomposed into simpler disjoint FTs, which are "analyzable". The results from the analysis of all simpler FTs are re-combined to obtain the results for the original FT. Using the presented method, it's possible to obtain minimal cut sets and the probability of the failure of the system. However, it should be noted that, because of decomposing and re-combining the FT, the determined value of the failure of the system will contain an error.

Paper [9] focuses on a sub-class of dynamic FT, which are called Priority Dynamic FTs. Described Priority Dynamic FTs use so called Priority Dynamic Gates and Repeated Events to consider dependencies between events, including failure events. However, further analysis of such FTs requires the use of SM methods or state space method.

Paper [10] investigates the problem of mutual influence of basic events in Noncoherent FTs. Authors present an approach of the development of FTs, which considers changing the failure probability at certain conditions. The results are demonstrated with an example, where authors develop the model of emergency shutdown system in nuclear power plant. This approach is complicated, involves significant effort and engineer time for FT design and analysis.

Article [11] compares two methods of analysis – FT analysis and failure mode and effects analysis (FMEA). Authors present a comparative analysis of these two tools, showing the contribution of each one for the implementation of a structured predictive maintenance planning for the hydraulic turbines.

Work [12] presents method of the analysis of the reliability of telecommunication systems. Presented method is based on FT analysis. As a part of this method, author proposes principles of the development of models of telecommunication systems together with models of their technical exploitation systems. Such models can be used for ensuring the reliability of systems and for determining the importance of failures of the system. With the development of the technical exploitation models, author emphasize the importance of considering not only the structure of the technical system, but also its functional behavior.

Authors of an article [13] propose the method of FT development, which is based on load-sharing redundancy. The main disadvantage of the proposed method is that is difficult to use and it requires large amount of time for model development.

The overview of modern methods of FT development and analysis shows us, that in most cases, "manual work" is required. Also, most methods consider that all basic events in the system are independent (FT is static). On the other hand, use of dynamic FTs, which allows to consider interconnection between events, requires involving complicated methods for further analysis (SM, states space method), that entails significant costs of time and effort.

Besides that, known methods of FTs development and analysis are not able to consider reliability and functional behavior of the CTS in one model. The maintenance and limitations of repairs number, the tools of control and diagnosis, the reliability of software are not considered also.

Therefore, there is an important problem of development of an approach, which would allow to consider aforementioned features of CTS. It's also important for such approach to be well-formalized so it would be possible to automate the process of obtaining FT and MCS.

III. The proposed approach for the automation of the development of the fault tree of the complex technical system

The proposed approach for the automation of the development of FTs is based on an existing method of determining MCS directly from the state-transition model of the object of investigation [14]. The other basis of the proposed approach is the state space method [15], in particular its improved version [16, 17]. This improved version of the state space method is based on the well-formalized method of the development of the state-transition models, which allows to automate the development of mathematical models of the behavior of CTS, considered as discrete-continuous stochastic systems.

Fig. 1. Fault tree development methodology

The automation of the development of mathematical models of the behavior is based on so-called "structural-automaton" model (SAM) of the object of investigation. SAM of CTS is a formalized representation of its structure and behavior. SAM consists of three parts. First part of it is a set of parameters, considered in the model. These parameters are listed in a state vector. Second part – list of events that cause change in the system's state. These events are called "basic events". Third part of SAM – set of expressions that describe the rules of state vector modification, caused by basic events. There is an algorithm to develop state-transition model from SAM. However, the level of detail of representation of the object of investigation in obtained model is insufficient for developing the fault tree (FT). For solving this problem, the method of development of the SAM has been improved so we could develop so called "binary SAM".

The binary SAM is the SAM of the object of investigation, in which every considered component of system's structure is represented by individual state vector component which can take only zero/one values. The point of developing the binary SAM, is that, unlike to ordinary SAM [15, 16], in binary SAM it is possible to describe the structure and behavior of CTS without coupling states of its structure components. Thereby, it is possible to obtain the

description of specific failure state which will include states of inoperative components with required level of details.

Using ASNA software tool [17], it is possible to generate state-transition model from binary SAM. State-transition model, generated from binary SAM, will have a sufficient level of details of the object of investigation, required for FT development. For generated state-transitions model, ASNA software tool builds the analytical model of CTS in form of the system of linear differential Chapman - Kolmogorov equations. The solution of these equations gives the probability distribution of the CTS being in all possible states. Using the proposed procedure of filtering failure states, MCSs can be obtained [14]. Quantitative indicators of MCSs can be obtained by summing the probabilities of the system being in corresponding states.

Described technique of FT development is presented on Figure 1.

IV. THE AUTOMATION OF THE DEVELOPMENT OF THE FAULT TREE OF COMPLEX TECHNICAL SYSTEM

In proposed approach, the MCS array is used as an input data for FT development. As a result of procedures of the proposed approach, the logical function will be obtained. With this logical function, the graphical representation of FT can be built.

Automated FT development is performed in two stages. On the first stage FT logical function is obtained by sequential sorting of MCS array. Based on obtained FT logical function the FT graphical representation is built on the second stage.

It should be noted that obtained FT is static but probabilistic indexes of MCS consider dynamic processes which occur in CTS.

According to presented technique an algorithm which builds FT logical function is developed and is implemented in software prototype.

A. Algorithm of obtaining the fault tree logical function

In presented algorithm of the first stage the sequential record of FT logical function is performed.

Abbreviations which are used in description of algorithm for automated obtaining of FT logical function:

n – MCS serial number pointer.

i – serial number of the SV component pointer.

j – zero value of SV component counter; this counter keeps the number of SV components whose values are zero.

CSVC – constant of SV components number.

ZVC_n – zero value counter; this counter keeps the number of SV components whose values are zero for MCS with serial number n.

CMCS – constant of MCS number.

$SV_n[i]$ – value of the i-th SV component for MCS with serial number n.

LFF – logical function form.

Step 1. Take the first MCS. In MCS serial number pointer n the 1 is recorded and the number of SV components, whose values are zero for the first MCS, is recorded in zero value counter ZVC_1 of first MCS. In zero value SV component counter j the number of SV components, whose values are zero, is recorded: $j = ZVC_1$.

Step 2. The first SV component of the first MCS is taken: $i = 1.$ In LFF a "(" symbol is written.

Step 3. On this step the value of i-th SV component of the MCS with serial number n is needed to be determined (in first cycle of algorithm the first SV component of the first MCS). So there such variants can take place:

If value of the **i**-th SV component of the MCS with serial number **n** is equal to zero ($SV_n[i]$ = 0), then in logical function form **(LFF)** the value of serial number of the SV component pointer **i** is written V**i**. Written zero value SV component counter **j** is decremented by one: **j = j-1** and after this **j** is checked by inequality **j < 0**.

If **j > 0** it means that in MCS with serial number **n** there are SV components whose values are zero. So in **LFF** a "·" symbol is written, then go to step 4.

If **j = 0** it means that in MCS with serial number **n** there are no more SV components whose values are zero, then go to step 4.

If value of the **i**-th SV component of the MCS with serial number **n** is not equal to zero ($SV_n[i]$ > 0), then go to step 4.

Step 4. Serial number of the SV component pointer **i** of the MCS with serial number **n** is incremented by one: **i = i+1.** After incrementation pointer **i** is checked by comparison with **CSVC.**

If **i ≤ CSVC** it means that the pointer **i** is still in array of SV components.

If **i > CSVC** it means that the pointer **i** got out last SV component:

If **i ≤ CSVC**, then go to step 3.

If **i > CSVC**, then go to step 5.

Step 5. Minimal cut sets serial number pointer **n** is incremented by one: **n = n+1**. Now it is checked if the pointer **n** got out MCS array by its comparison with **CMCS**:

If **n ≤ CMCS**, then in **LFF** the ") +" symbols are written and go to step 2.

If **n > CMCS**, then in **LFF** the ");" symbols are written. In this stage the procedure of writing FT logical function is finished.

B. *Algorithm of building the fault tree graphical representation*

Step 1. First stage is the building of the lowest levels of FT. For this all elements, which are in brackets and between which are multiply symbols, are combined by logical elements AND. All groups of elements, between whose elements there are multiply symbols, have only one output. In result the lowest level of fault tree is built. After that it should go to step 2.

Step 2. In this stage next levels and the highest event of fault tree are formed. For this all outputs from lower levels are combined by logical elements OR according to elements combining rule.

Elements combining rule – logical elements AND or OR can combine only two outputs from lower level or two elements in step 1.

Using this technique, the FT is built.

V. FAULT TREE BUILDING EXAMPLE OF FAULT TOLERANT HARDWARE/SOFTWARE SYSTEM

Fault-tolerant hardware/software system (FTH/SS) is object which consists of four main modules: input receiving module, operating module, data processing module and executing output module. All these modules perform an objective function of FTH/SS. Input receiving module and data processing module are reserved. Failure of FTH/SS can occur in result of hardware or software failure. Block diagram of FTH/SS is shown in Figure 2.

According to presented in [14] technique the MCS array was obtained for FTH/SS and it is presented in Table 1.

Input data:

CSVC – assigns to 7 – there are 7 SV components.

CMCS – assigns to 4 – there are 4 MCS.

Fig.2. Block diagram of fault-tolerant
hardware/software system.

TABLE I
MINIMAL CUT SETS ARRAY

Serial numbers of MCS	State vector components
1	V1=1; V2=1; V3=1; V4=1; V5=1; V6=1; V7=0
2	V1=1; V2=1; V3=1; V4=1; V5=0; V6=0; V7=1
3	V1=1; V2=1; V3=0; V4=0; V5=1; V6=1; V7=1
4	V1=0; V2=0; V3=1; V4=1; V5=1; V6=1; V7=1

A. Algorithm of obtaining the fault tree logical function

Step 1. Take the **1**-st MCS and the **1**-st SV component: **n = 1**; **j = 1**;

Step 2. **i** pointer assigns to 1: **i = 1;**

In **LFF** write a "(" symbol and go to step 3.

Step 3. Take value of the **1**-st SV component of the **1**-st MCS. As **(SV$_1$[1] = 1)** > 0, so go to step 4.

Step 4. **i = i+1; i = 1+1=2.** As **2 < 7 (i ≤ CSVC)**, so go to step 3.

Step 3. Take value of the **2**-nd SV component of the **1**-st MCS. As **(SV$_1$[2] = 1)** > 0, so go to step 4.

Step 4. **i = i+1; i = 2+1=3.** As **3 < 7 (i ≤ CSVC)**, so go to step 3.

Step 3. Take value of the **3**-rd SV component of the **1**-st MCS. As **(SV$_1$[3] = 1)** > 0, so go to step 4.

Step 4. **i = i+1; i = 3+1 = 4.** As **4 < 7(i ≤ CSVC)**, so go to step 3.

Step 3. Take value of the **4**-th SV component of the **1**-st MCS. As **(SV$_1$[4] = 1)** > 0, so go to step 4.

Step 4. **i = i+1; i = 4+1 = 5.** As **5 < 7 (i ≤ CSVC)**, so go to step 3.

Step 3. Take value of the **5**-th SV component of the **1**-st MCS. As **(SV$_1$[5] = 1)** > 0, so go to step 4.

Step 4. **i = i+1; i = 5+1 = 6.** As **6 < 7 (i ≤ CSVC)**, so go to step 3.

Step 3. Take value of the **6**-th SV component of the **1**-st MCS. As **(SV$_1$[6] = 1)** > 0, so go to step 4.

Step 4. **i = i+1; i = 6+1 = 7.** As **7 = 7 (i ≤ CSVC)**, so go to step 3.

Step 3. Take value of the 7-th SV component of the **1**-st MCS. **SV$_1$[7] = 0**.

In **LFF** write V**i** – **V7**.

j = j-1; j = 1-1 = 0. As **j = 0**, so go to step 4.

LFF: (V7

Step 4. **i = i+1; i = 7+1 = 8.** As **8 > 7 (i > CSVC)**, so go to step 5.

Step 5. **n = n+1; n = 1+1=2.** As **2 < 4 (n ≤ CMCS)**, so in **LFF** write ") +" symbols and go to step 2.

LFF: (V7) +

Step 2. **i = 1**; **j = 2**. In **LFF** write "(" symbol and go to step 3:

LFF: (V7) + (

Step 3. Take value of the **1**-st SV component of the **2**-nd MCS. As **(SV$_2$[1] = 1)** > 0, so go to step 4.

Step 4. **i = i+1; i = 1+1 = 2.** As **2 < 7 (i ≤ CSVC)**, sogo to step 3.

Step 3. Take value of the **2**-nd SV component of the **2**-nd MCS. As **(SV$_2$[2] = 1)** > 0, so go to step 4.

Step 4. **i = i+1; i = 2+1 = 3;** As **3 < 7 (i ≤ CSVC)**, so go to 3.

Step 3. Take value of the **3**-rd SV component of the **2**-nd MCS. As **(SV$_2$[3] = 1)** > 0, so go to step 4.

Step 4. **i = i+1; i = 3+1 = 4. 4 < 7 (i ≤ CSVC)**, go to step 3.

Step 3. Take value of the **4**-th SV component of the **2**-nd MCS. As **(SV$_2$[4] = 1)** > 0, so go to step 4.

Step 4. **i = i+1; i = 4+1 = 5.** As **5 < 7 (i ≤ CSVC)**, so go to step 3.

Step 3. Take value of the **5**-th SV component of the **2**-nd MCS. **SV$_2$[5] = 0**.

In **LFF** write V**i** – V5:

j = j -1; j = 2-1 = 1. As **j > 0**, so write a "·" symbol and go to step 4.

LFF: (V7) + (V5 ·

Step 4. **i = i+1; i = 5+1 = 6.** As **6 < 7 (i ≤ CSVC)**, so go to step 3.

Step 3. Take value of the **6**-th SV component of the **2**-nd MCS. **SV$_2$[6] = 0.**

In **LFF** write V**i** – V6:

LFF: (V7) + (V5 · V6

j = j-1; j = 1-1=0; As **j = 0**, so go to step 4.

Step 4. **i = i+1; i = 6+1 = 7.** As **7 = 7 (i ≤ CSVC)**, so go to step 3.

Step 3. Take value of the **7**-th SV component of the **2**-nd MCS. As **(SV$_2$[7] = 1)** > 0, so go to step 4.

Step 4. **i = i+1; i = 7+1 = 8.** As **i > CSVC**, so go to step 5.

Step 5. **n = n+1; n = 2+1 = 3;** As **3≤ 4 (n < CMCS)**, so in **LFF** write ") +" symbols and go to step 2.

LFF: (V7) + (V5 · V6) +

Step 2. **i = 1; j = 2.**

In **LFF** write a "(" symbol and go to step 3:

LFF: (V7) + (V5· V6) + (

Step 3. Take value of the **1**-st SV component of the **3**-rd MCS. As **(SV$_3$[1] = 1)** > 0, so go to step 4.

Step 4. **i = i+1; i = 1+1 = 2.** As **2 < 7 (i ≤ CSVC)**, so go to step 3.

Step 3. Take value of the **2**-nd SV component of the **3**-rd MCS. As **(SV$_3$[2] = 1)** > 0, so go to step 4.

Step 4. **i = i+1; i = 2+1 = 3.** As **3 < 7(i ≤ CSVC)**, so go to step 3.

Step 3. **SV$_3$[3] = 0.**

In **LFF** write V**i** – V3:

LFF: (V7) + (V5· V6) + (V3

j = j-1; j = 2-1=1. As **j > 0** so write a "·" symbol and go to step 4.

LFF: (V7) + (V5· V6) + (V3 ·

Step 4. **i = i+1; i = 3+1=4.** As **4 < 7 (i ≤ CSVC)**, so go to step 3.

Step 3. **SV$_3$[4] = 0;**

In **LFF** write V**i** – V4:

LFF: (V7) + (V5· V6) + (V3 · V4

j = j -1; j = 2 -1=1. As **j = 0,** so go to step 4.

Step 4. **i = i+1; i = 4+1=5;** As **5 < 7 (i ≤ CSVC),** so go to step 3.

Step 3. Take value of the **5**-th SV component of the **3**-rd MCS. As **(SV₃[5] = 1)** > 0, so go to step 4.

Step 4. **i = i+1; i = 5+1=6.** As **6 < 7 (i ≤ CSVC)**, so go to step 3.

Step 3. Take value of the **6**-th SV component of the **3**-rd MCS. As **(SV₃[6] = 1)** > 0, so go to step 4.

Step 4. **i = i+1; i = 6+1=7;** As **7 = 7 (i ≤ CSVC)**, so go to step 3.

Step 3. Take value of the **7**-th SV component of the **3**-rd MCS**.** As **(SV₃[7] = 1)** > 0, so go to step 4.

Step 4. **i = i+1; i = 7+1=8.** As **8 > 7 (i > CSVC)**, so go to step 5.

Step 5. **n = n+1; n = 3+1= 4.** As **4 = 4 (n ≤ CMCS)**, so in **LFF** write ") +" symbols and go to step 2.

LFF: (V7) + (V5 · V6) + (V3 · V4) +

Step 2. **i = 1; j = 2.**

In **LFF** write a "(" symbol and go to step 3:

LFF: (V7) + (V5 · V6) + (V3 · V4) + (

Step 3. Take value of the **1**-st SV component of the **4**-th MCS**. SV₄[1] = 0**.

In **LFF** write V**i** – V1.

j = j-1; j = 2-1=1. As **j > 0,** so write a "·" symbol and go to step 4.

LFF: (V7) + (V5 · V6) + (V3 · V4) + (V1 ·

Step 4. **i = i+1; i = 1+1=2.** As **2 < 7 (i ≤ CSVC),** so go to step 3.

Step 3. Take value of the **2**-nd SV component of **4**-th MCS. **SV₄[2] = 0.**

In **LFF** write V**i** – V2.

LFF: (V7) + (V5 · V6) + (V3 · V4) + (V1 · V2

j = j-1; j = 1-1=0. As **j = 0**, so go to step 4.

Step 4. **i = i+1; i = 2+1=3.** As **3 < 7 (i ≤ CSVC)**, so go to step 3.

Step 3. Take value of the **3**-rd SV component of **4**-th MCS. As **(SV₄[3] = 1)** > 0, so go to step 4.

Step 4. **i = i+1; i = 3+1=4.** As **4 < 7 (i ≤ CSVC)**, so go to step 3.

Step 3. Take value of the **4**-th SV component of **4**-th MCS. As **(SV₄[4] = 1)** > 0, so go to step 4.

Step 4. **i = i+1; i = 4+1=5.** As **5 < 7 (i ≤ CSVC)**, so go to step 3.

Step 3. Take value of the **5**-th SV component of the **4**-th MCS. As **(SV₄[5] = 1)** > 0, so go to step 4.

Step 4. **i = i+1; i = 5+1=6.** **6 < 7 (i ≤ CSVC)**, so go to step 3.

Step 3. Take value of the **6**-th SV component of the **4**-th MCS. **(SV₄[6] = 1)** > 0.

Step 4. **i = i+1; i = 6+1=7.** As **7 = 7 (i ≤ CSVC)**, so go to step 3.

Step 3. Take value of the **7**-th SV component of the **4**-th MCS. As **(SV₄[7] = 1)** > 0, so go to step 4.

Step 4. **i = i+1; i = 7+1=8.** As **8 > 7 (i > CSVC)**, so go to step 5.

Step 5. **n = n+1; n = 4+1=5;** As **5 > 4 (n > CMCS)**, so in **LFF** write ");" symbols.

LFF: (V7) + (V5 · V6) + (V3 · V4) + (V1 · V2);

Procedure of writing the FT logical representation is finished.

According to algorithm of building the FT graphical representation and basing on obtained LFF the FT of fault-tolerant H/S system was built and is presented in Figure 3.
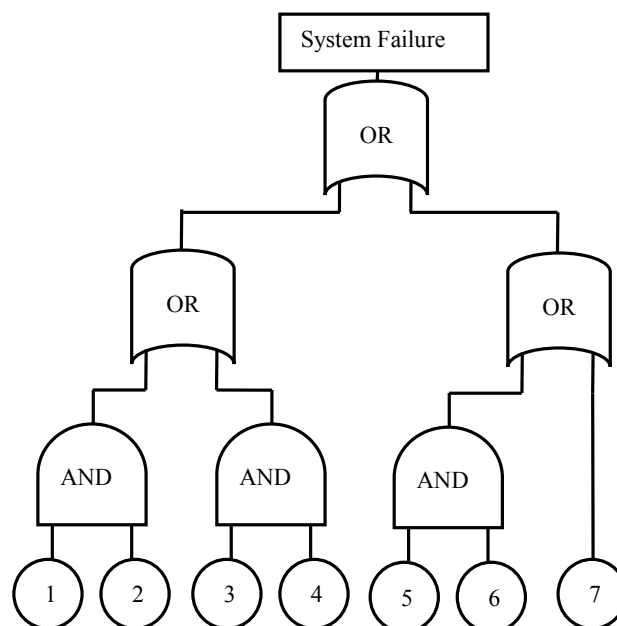
Fig.3. Fault tree of fault tolerant H/S

VI. CONCLUSIONS

The developed software tool that is based on presented approach significantly reduces time required for the analysis of reliability and exploitation safety of complex technical systems. Especially this reduction is noticeable when it is necessary to consider multiple configurations of investigated system, because the development of the fault tree for each implementation variant is automated.

Presented technique of automated fault tree development and analysis provides high certainty of recommendations regarding increasing reliability and exploitation safety of complex technical systems.

REFERENCES

[1] Mozhaev A. S. Theoretical Fundamentals, Experience of Application and Direction of Development of General Logical-probabilistic Method of Software Complex "ARBITR" for Reliability Modeling, Durability, Safety and System risk / A. S. Mozhaev // Proceedings of JSC "SPIK SZMA", St. Petersburg.- 2010.- 11 p. [in Russian]
[2] Shcherbovskykh S.V. Mathematical Models and Methods for Dependability Characteristic Determination of Repairable Multi-terminal Load-shared Systems / S.V. Shcherbovskykh. – Lviv: Lviv Polytechnic National University Publ., 2012. – 296 p. [in Ukrainian]
[3] Henley E. J. Reliability Engineering and Risk Assessment / E. J. Henley, H. Kumamoto // Prentice-Hall, - 1981. - 568 p.
[4] Kececioglu D. Reliability Engineering Handbook, Vol. 2 / D. Kececioglu, - Prentice Hall Inc.: New Jersey, 1991. - 541 p.
[5] Liggesmeyer P. Fault Tree Analysis, Current Research Issues, Tutorial / P. Liggesmeyer, B. Kaiser // Proceed SAFECOMP 2004, Potsdam 2004.
[6] Chiacchio F. Dynamic Fault Trees Resolution: a Conscious Trade-Off Between Analytical and Simulative Approaches / F. Chiacchio, L.Compagno, D. D'Urso, G. Manno, N.Trapani // Reliability Engineering & System Safety. - 2011. - Vol. 96, № 11. - P. 1515-1526.
[7] Skrobanek P. Analysis of Timing Requirements for Intrusion Detection and Prevention Using Fault Tree with Time Dependencies / P. Skrobanek, M.Woda // In: Skrobanek P, editor. Intrusion detection systems. InTech.- 2011.- P. 307-324.
[8] Contini S. Analysis of Large Fault Trees Based on Functional Decomposition / Sergio Contini, Vaidas Matuzas // Reliability Engineering & System Safety. - 2011. - Vol.96, № 3. - P. 383-390.

[9] Merle G. Probabilistic Algebraic Analysis of Fault Trees With Priority Dynamic Gates and Repeated Events / G. Merle, J.-M. Roussel, J.-J. Lesage, A. Bobbio // Reliability, IEEE Trans. on. - 2010. - Vol. 59, № 1. - P. 250-261.

[10] Lixuan Lu. Joint Failure Importance for Noncoherent Fault Trees / Lu Lixuan, Jin Jiang // Reliability, IEEE Trans. on. - 2007. - Vol. 56, № 3. - P. 435-443.

[11] Rodrigo de Queiroz Souza FMEA and FTA Analysis For Application of The Reliability Centered Maintenance Methodology: Case Study on Hydraulic Turbines / Rodrigo de Queiroz Souza, Alberto José Álvares // ABCM Symposium Series in Mechatronics - Vol. 3 - P. 803-812.

[12] Maevskyy L. S. Methods of Reliability Providing of Information - Telecommunication Systems in Different Life Cycle Stages / L. S. Maevskyy. - St. Petersburg: Barzilovich Z. P. Publ., 1999, 112 p. [in Russian]

[13] Mandziy B. Mathematical Model for Failure Cause Analysis of Electrical Systems with Load-sharing Redundancy   of   Component   / B. Mandziy, O. Lozynsky, S. Shcherbovskykh, // Przegląd Elektrotechniczny, 89 (2013), No 11, pp. 244-247.

[14] Volochiy, B. The new method of building a safety model for quantitative risk assessment of complex technical systems for critical application /Volochiy, B.,Mandziy, B.,Ozirkovskyy, L.// Communications in Computer and Information Science. Volume 594. Information and Communication Technologies in Education, Research, and Industrial Applications, Publisher: Springer International Publishing, 2016, pp.56-70

[15] Tarakanov K. V. Analytical Methods of Systems Research / K. V. Tarakanov, L. A. Ovcharov, A. N. Tyryshkin,- M .: Sov. Radio, 1974, 240 p. [in Russian]

[16] Volochiy B. Yu. Modeling Technology of Behavior Algorithms of Information Systems / Volochiy B.Yu. – Lviv: Lviv Polytechnic National University Publ., 2004, 220 p. [in Ukrainian]

[17] Bobalo Yu. Mathematical Models and Methods of Analysis of Radioelectronic, Electromechanic and Software Systems / Yu. Bobalo, B. Volochiy, O. Lozynskyy, B. Mandziy, L. Ozirkovskyy, D. Fedasyuk, S. Shcherbovskykh , V. Yakovyna. – Lviv: Lviv Polytechnic National University Publ., 2013, 300 p. [in Ukrainian]

Parser of Input Data in Reliability Analysis based on Logical Differential Calculus

Patrik Rusnak

Abstract—One of the principal data that are used as an input of many algorithms in reliability analysis is structure function, which defines the correlation between the system performance and performance of its components. It has been shown in several papers that this function can also be viewed as a Multiple-Valued Logic (MVL) function. This idea allows us to use methods related to the investigation of MVL functions in reliability analysis. One of them is logical differential calculus, which can be used to find circumstances under which degradation of a specific system component results in a decrease in system operation. MVL functions can be represented in several ways, e.g., truth table, graphic form, symbolic form. Computer processing of MVL functions requires a specific parser that is able to transform a given representation of a MVL function into a form that can be easily processed on the computer. In this paper, the symbolic representation is considered primarily. Parsing symbolic expressions can be done using several universal algorithms. One of them is shunting-yard algorithm invented by Edsger Dijkstra. Implementation of this algorithm for parsing MVL functions but also general mathematical expressions is presented in this paper.

Keywords—reliability, logical differential calculus, shunting-yard algorithm.

I. INTRODUCTION

Reliability is an important characteristic of many, not only technical, systems. One of the current issues of reliability analysis is investigation of complex systems [1]. Such systems are composed of many components that are very different in their behavior. Typical examples of complex systems are healthcare systems, which consist of components of different nature that can be classified as hardware, software, and human factor [2], or distribution networks, which are composed of hardware elements with very different behavior [3]. Investigation of such systems requires development of new methods that take this diversity into account. One of the possible ways of how this diversity can be modeled is application of multi-state models.

A multi-state model of a system allows defining several performance levels at which the system or its components can operate. These levels are known as system/components states. A map that defines the dependency between components states and state of the system is known as structure function, and it has the following form [4]:

$$\phi(x_1, x_2, \dots, x_n) = \phi(\pmb{x})\colon \; \{0, 1, \dots, m-1\}^n \to \{0, 1, \dots, m-1\}, \tag{1}$$

where n denotes number of system components, m agrees with the number of system/ components states (where state 0 means complete failure while $m-1$ agrees with perfect functioning), x_i is a variable representing state of the i-th system component, and $\pmb{x} = (x_1, x_2, \dots, x_n)$ is a vector of components states (state vector).

The definition of the structure function agrees with the formal definition of Multiple-Valued Logic (MVL) function. This fact allows using some tools related to the analysis of MVL functions in reliability analysis of systems modeled using multi-state approach [5]. One of these tools is logical differential calculus.

Logical differential calculus has been developed for investigation of dynamic properties of MVL functions. Its central term is logic derivative. Several types of logic derivatives exists [6] but, in reliability analysis, the most useful one is Direct Partial Logic Derivative (DPLD). With

P. Rusnak, University of Zilina, Faculty of Management Science and Informatics, Zilina, Slovakia (e-mail: pattrik.rusnak@gmail.com).

respect to MVL function $\phi(x)$, this derivative is defined as follows [6]:

$$\partial\phi(j \rightarrow h)/\partial x_i(s \rightarrow r) = \begin{cases} 1, & \text{if } \phi(s_i, x) = j \text{ AND } \phi(r_i, x) = h \\ 0, & \text{otherwise} \end{cases},$$ (2)

$$\text{for } s, r, j, h \in \{0, 1, \dots, m-1\}, s \neq r, j \neq h,$$

where $(a_i, x) = (x_1, x_2, \dots, x_{i-1}, a, x_{i+1}, \dots, x_n)$ for $a \in \{s, r\}$. As we can see, a DPLD allows identifying situations in which a given change of a MVL variable results in a given change of the investigated MVL function. In reliability analysis, this permits finding state vectors at which degradation (improvement) of component i from state s to r results in degradation (improvement) of system state from j to h. Knowledge of such state vectors plays a key role in many fields of reliability analysis because it allows evaluating influence of a considered component on system operation, what can be used in optimization of system reliability or in planning system maintenance. Theoretical background for these ideas has been developed in several works. In [5, 7], it has been shown how DPLDs can be used to investigate importance of individual system components or their states. Works [8, 9] presented application of logical differential calculus in finding minimal scenarios whose occurrence results in system degradation (improvement).

The aforementioned papers have introduced a complex framework for reliability analysis based on logical differential calculus. However, the detailed computer implementation of that framework has not yet been considered. This problem is considered in this paper. Our goal is to develop a complex tool that will implement all of the methods proposed in the previously mentioned papers. The tool has to be efficient and universal, i.e. it has to be able to run efficiently on any type of input data. In our case, the input data is a structure function represented using a MVL function. MVL function can be expressed in many forms, e.g., tabular or symbolic forms. If we want to work with symbolic forms, then we need a parser that will be able to transform symbolic expression in the form that can be processed on a computer. One of the possible solutions to this problem is use of shunting-yard algorithm invented by Edsger Dijkstra [10, 11]. Practical implementation of this algorithm is considered in the rest of the paper.

II. SHUNTING-YARD ALGORITHM

The main principle of the shunting-yard algorithm is to process mathematical expressions specified in infix notation (e.g.: $2^3 + 4 * 5$) to the form of a reverse polish notation (i.e., $23 \wedge 45 * +$) [12, 13]. In our implementation of this algorithm, we will use two stacks: the first stack named output stack will store the output set of tokens, the second stack named temporary stack will store functions, operators, parentheses, and function arguments separators to maintain the priority of every operation.

Token is a pair consisting of a token name and an optional attribute value. The token name is an abstract symbol representing a kind of lexical unit, e.g., a particular keyword, or a sequence of input characters denoting an identifier. The token names are the input symbols that the parser processes. We will often refer to a token by its token name [12].

The stack used in the shunting-yard algorithm is an abstract data type, which is a collection of elements with two principal operations: push, which adds an element to the collection, and pop, which removes the last element that has not been removed. Additionally, a peek operation may be defined, which gives access to the first element of the stack without modifying it. The order in which elements are removed from the stack (last in, first out) is the basis for its alternative name LIFO [14].

The algorithm will recognize all tokens in input text and insert them into one of the two stacks by following these rules:

- If the token is a number or constant, then push it onto the output stack.
- If the token is an operation, then push it onto the temporary stack. But before that happens, you need to check the priority of the operation on the top. If on the top of the temporary stack is operation with higher or equal priority and inserting operation is left associative or when inserting operation is right associative and on the top of the temporary stack is operation with higher priority, then the operation at the top is pulled from the temporary stack, and it is pushed to the output stack. This checking process is repeated, until checking condition is no longer valid or the temporary stack is empty.
- If the token is a function token, then push it onto the output stack. However, when token is a function parameters separator, then tokens from the temporary stack are popped and pushed to the output stack until at the top of the temporary stack is the beginning of the function. If the beginning of the function is not found, then the function and the input text are invalid.
- If the token is a left parenthesis, then push it onto the output stack. But if the token is a right parenthesis, then tokens from the temporary stack are popped and pushed to the output stack until at the top of the temporary stack is a left parenthesis. Pop the left parenthesis from the temporary stack, but not onto the output stack. If the temporary stack runs out without finding a left parenthesis, then there are mismatched parentheses.
- When there are no more tokens to read, pop all tokens from the temporary stack and push them to the output stack.

An illustrative example of parsing input text using the shunting-yard algorithm can be seen in Fig. 1 where input string $2 \wedge 3 + 4 * 5$ is parsed.

Fig. 1 Illustration of the shunting-yard algorithm

III. BUILDING MULTI-WAY TREE

After we successfully execute shunting-yard algorithm, we get all the tokens in reverse polish notation in the output stack. But if we want to perform some specific action with output stack, e.g., symbolic computations such as simplification, then it will be difficult. We need to figure out a better way to keep output tokens and take advantage from the output stack. We concluded that the best solution will be in building multi-way tree from the output stack.

A tree is an abstract data type or data structure that implements this abstract data type. A tree data structure can be defined recursively as a collection of nodes (starting at a root node), where each node is a data structure consisting of a value, together with a list of references to nodes (the "children"), with the constraints that no reference is duplicated, and none points to the root [14].

Building a multi-way tree is performed by algorithm, which illustration can be noticed in Fig. 2. Firstly, we must ensure that the output stack is not empty. If it is not empty, then we can begin the following algorithm:

1. Create a node and set this node as the root.
2. Call this this recursive operation that will operate as follows:
    2.1. Select a token from the output stack.
    2.2. If the token does not have arguments, then end.
    2.3. If the token has n arguments, then repeat these steps n times:
        2.3.1. Create a new node.
        2.3.2. Set the node as a son of the token.
        2.3.3. Run recursive operation over the node.

After the algorithm run, we get a multi-way tree, which will run demanding tasks much simpler and more practical.

Fig. 2 Building a multi-way tree

IV. IMPLEMENTATION OF SHUNTING-YARD ALGORITHM

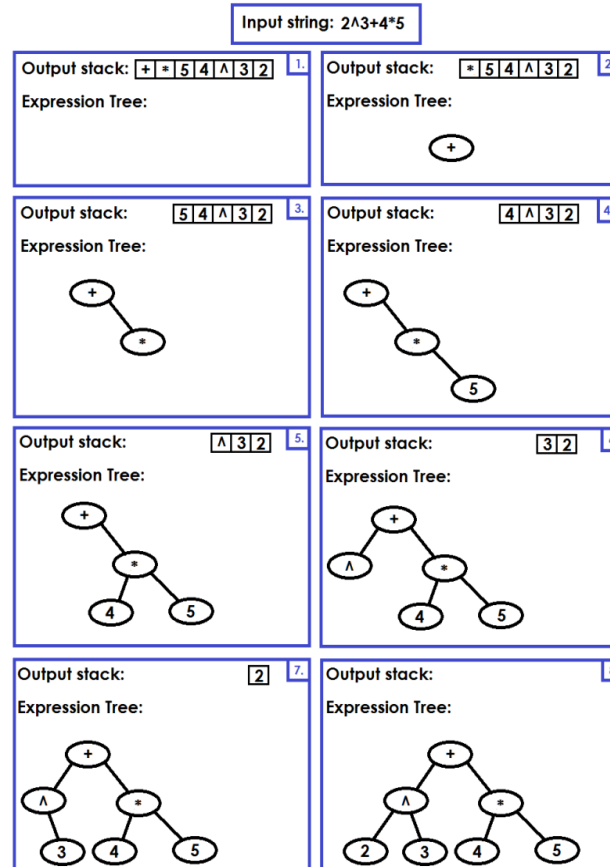In order to implement the shunting-yard algorithm, building a multi-way tree and ensure calculation for complex framework for reliability analysis based on logical differential calculus, we have to think of the class architecture for the developed tool. Basic module, responsible for parsing and evaluating expressions (can be seen in Fig. 3) has been designed so that it can process the input expression and convert it into the shape of the multi-way tree. The class diagram consists of 10 classes, whose individual meaning will now be explained.
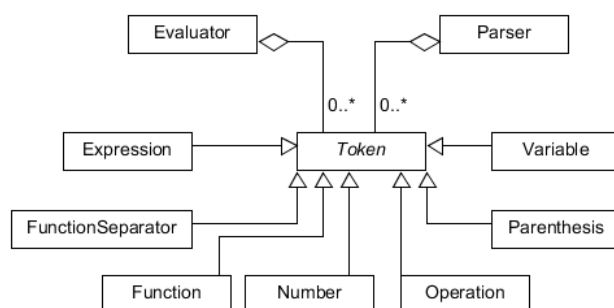

Fig. 3 Class diagram

The first one is abstract class Token, which is a superclass for any supported type of expression, such as constants, operators, functions, etc. Variable class, a subclass of class Token, represents a variable and this variable has to meet certain established rules such as variable must begin with capital or small letter, $ or break character; the rest characters must be capital or small letters, digits or break characters and variables could not be named as defined functions or operations (e.g., X1, variable_one, ThisIsAlsoVariable). Number is a class (a subclass of class Token), which represents the numeric value or defined constants. The occurrence of parentheses and their impact on the priority has also to be taken into account during parsing, so Parenthesis class, a subclass of class Token, is needed for their representation.

The previous classes allow us to use variables, constants and parentheses in the program. Another important part is support for operations and functions. For this purpose, the next classes are defined in the parser.

Operation class, a subclass of class Token, represents all operations in the specified system. This class stores representation in the input, operation's priority, associativity and arity. Another important class is Function class, which is also a subclass of class Token. This class presents a general representation of the functions defined in the system. Functions can have multiple parameters. These parameters are separated by function parameters separator. In order to take function parameters separator into account when parsing, it was necessary to define a class FuncionSeparator, which is also inherit from abstract class Token.

Sometimes, it is also necessary to represent the expression, e.g. $5 * x + 6$. For this purpose, the Expression class can be used.

One of the main classes in the module is the class Parser, which executes shunting-yard algorithm based on the specified input text and stores the output stack for further usage. It also allows us to set the characters used in the input text, such as decimal point, types of parentheses and function parameters separator.

The other one of the main classes is class Evaluator. An instance of this class is responsible for creating a multi-way tree based on the stack received from an instance of class Parser. After creating the tree, instances of this class allow us to perform all symbolic calculations on the symbolic expression, such as simplification or, in case of logic functions, transformation into normal forms.

V. CONCLUSION

One of the current issues of reliability analysis is investigation of complex systems. Several approaches can be used in solving this problem. One of them is based on application of MVL. MVL functions and tools related to them (e.g., logical differential calculus) can be used to express the structure function of a system under consideration, investigate importance of the system components, or find minimal scenarios needed for ensuring system mission. Application of these ideas in the analysis of systems composed of many components requires creating complex software that will implement all of them. However, such software has not yet been developed. Because of that, we decided to create one.

The essential part of the previously mentioned software is a module that will be able to parse and transform complex mathematical expressions, which are easily readable by a human, into the form that can be processed on the computer. For this purpose, we decided to implement the shunting-yard algorithm and use a multi-way tree to represent the parsed expression. In this paper, we present the architecture of the module that implements this algorithm. The architecture reflects all elements that can exist in the mathematical expressions (variables, constants, functions, operators and their properties), and it is fully customizable for specific, e.g. logic, expressions, i.e., it allows defining the format of variables, possible constants, and operators with their properties, such as precedence, associativity, and arity.

REFERENCES

[1] E. Zio, "Reliability engineering: Old problems and new challenges," *Reliability Engineering & System Safety*, vol. 94, no. 2, pp. 125–141, Feb. 2009.

[2] E. Zaitseva, V. Levashenko, and M. Rusin, "Reliability analysis of healthcare system," in *2011 Federated Conference on Computer Science and Information Systems, FedCSIS 2011*, 2011, pp. 169–175.

[3] P. Praks, V. Kopustinskas, and M. Masera, "Probabilistic modelling of security of supply in gas networks and evaluation of new infrastructure," *Reliability Engineering & System Safety*, vol. 144, pp. 254–264, Dec. 2015.

[4] B. Natvig, *Multistate Systems Reliability Theory with Applications*. Chichester, UK: John Wiley & Sons, Ltd, 2011.

[5] E. Zaitseva and V. Levashenko, "Multiple-valued logic mathematical approaches for multi-state system reliability analysis," *Journal of Applied Logic*, vol. 11, no. 3, pp. 350–362, Sep. 2013.

[6] S. N. Yanushkevich, D. M. Miller, V. P. Shmerko, and R. S. Stankovic, *Decision Diagram Techniques for Micro- and Nanoelectronic Design Handbook*, vol. 2. Boca Raton, FL: CRC Press, 2005.

[7] M. Kvassay, E. Zaitseva, J. Kostolny, and V. Levashenko, "Importance analysis of multi-state systems based on integrated direct partial logic derivatives," in *2015 International Conference on Information and Digital Technologies (IDT)*, 2015, pp. 183–195.

[8] M. Kvassay, E. Zaitseva, V. Levashenko, and J. Kostolny, "Minimal cut vectors and logical differential calculus," in *2014 IEEE 44th International Symposium on Multiple-Valued Logic*, 2014, pp. 167–172.

[9] M. Kvassay, E. Zaitseva, and V. Levashenko, "Minimal cut sets and direct partial logic derivatives in reliability analysis," in *Safety and Reliability: Methodology and Applications - Proceedings of the European Safety and Reliability Conference, ESREL 2014*, 2015, pp. 241–248.

[10] Edsger W. Dijkstra, "Algol 60 translation : An Algol 60 translator for the x1 and Making a translator for Algol 60" in *Research Report 35, Mathematisch Centrum, Amsterdam, 1961*. Reprint archived at http://www.cs.utexas.edu/users/EWD/MCReps/MR35.PDF.

[11] T. S. Norvell "Parsing Expressions by Recursive Descent," Spring 1999 http://www.engr.mun.ca/~theo/Misc/exp_parsing.htm.

[12] A. V. Aho, M. S. Lam, R. Sethi and J. D. Ullman, *Compilers: Principles, Techniques, & Tools, Second Edition*, Pearson Education, Inc., 2007.

[13] B. R. Preiss, *Data Structures and Algorithms with Object-Oriented Design Patterns in C++*, John Wiley & Sons, 1999.

[14] T. H. Cormen, C. E. Leiserson and R. L. Rivest, *Introduction to Algorithms (3rd ed.)*, MIT Press and McGraw-Hill, 2009.

A Technique for Outlier Detection Based on Heuristic Possibilistic Clustering

Dmitri A. Viattchenin

Abstract—The paper deals with the problem of outlier detection in the data set. A corresponding technique is proposed in the paper. The technique is based on the heuristic approach to possibilistic clustering. A brief review of basic concepts and algorithms of the heuristic approach to possibilistic clustering is provided. A general plan of the proposed technique for outlier detection in the data set is proposed. An illustrative example is given. Some preliminary conclusions are formulated.

Keywords—heuristic possibilistic clustering, outlier detection, principal allotment among fuzzy clusters, typical point.

I. Introduction

The first subsection of the introduction includes a consideration of a problem of outlier detection by using fuzzy clustering methods. Related works are considered briefly in the second subsection.

A. Preliminary Remarks

Fuzzy clustering is used when the boundaries among the clusters are uncertain and confusing. Fuzzy clustering is well established area, and fuzzy clustering algorithms are standard tools in unsupervised machine learning and applied statistics.

A possibilistic approach to clustering was proposed by Krishnapuram and Keller [1] and the approach can be considered as a special case of fuzzy approach to clustering because all methods of possibilistic clustering are objective function-based methods. On the other hand, constraints in the possibilistic approach to clustering are less strong than constraints in the fuzzy objective function-based approach to clustering and values of the membership function of a possibilistic partition can be considered as typicality degrees. So, the possibilistic approach to clustering is more general and flexible approach to clustering than the fuzzy approach. Many fuzzy and possibilistic clustering algorithms could be found in the corresponding books, for example, in [2].

One of the basic problems of data mining is the outlier detection [3]. Detecting the outliers of a data set is an important research way for data cleaning and finding new useful knowledge in many research areas. Outliers are objects, which deviate significantly from the rest of the data, so that it seems they are determined by strange process. Of course, outliers are often bad data points. On the other hand, in many applications outliers contain important information and their correct identification is crucial. An illustrative example is a computer security intrusion detection system, which finds outlier patterns as a possible intrusion attempts. Intrusion detection corresponds to a suite of techniques that are used to identify attacks against computers and network infrastructures. Anomaly detection is a key element of intrusion detection in which perturbations of normal behavior suggest the presence of intentionally or unintentionally induced attacks, faults and defects. So, identifying outliers is an important step in data mining.

The aim of the presented paper is a consideration of the problem of discovering outliers in the data. A novel technique for outlier detection is proposed in the paper. However, the previous results should be considered in the first place.

D. A. Viattchenin, National Academy of Sciences of Belarus, Minsk, Belarus (e-mail: viattchenin@mail.ru).

B. Related Works

Let us consider some fuzzy clustering-based approaches and techniques for outlier detection in the data. These approaches consider clusters of small sizes as clustered outliers. In other words, clusters which containing significantly less objects than other clusters should be considered as outliers.

In the first place, the fuzzy kernel clustering algorithm with outliers was proposed in [4]. An outstanding property of the developed FKCO-algorithm is that the FKCO-algorithm cannot obtain a satisfying clustering performance, but can identify the outliers easily with the newly defined criteria.

In the second place, robust version of the well-known maximum entropy clustering algorithm [5] is presented in [6]. This RMEC-algorithm is based on Vapnik's ε-intensive loss function [7] and its principal advantage is capability of labeling outliers in a data set using the obtained weighting factors. It should be noted, that Vapnik's ε-intensive loss function was used by Łęski for developing the εPCM-algorithm of possibilistic clustering [8].

In the fourth place, a method to estimate the noise distance in the noise clustering on the preservation of the hypervolume of the feature space was proposed by Rehm, Klawonn and Kruse in [9]. The proposed approach is independent of the number of fuzzy clusters in the data set. Moreover, this approach is not only to reduce the influence of outliers, but also to identify them.

In the fifth place, the robust interval competitive agglomeration clustering algorithm is described in [10]. This RICA-algorithm is developed to overcome the problems of outliers, the unknown clusters number and the initialization of prototypes in the clustering procedure for the interval-valued data.

In the sixth place, a method for outlier detection is proposed in [11]. The method is based on fuzzy clustering. In particular, the FCM-algorithm is first performed, and then small clusters are detected and considered as outlier clusters. Other outliers are then detected based on computing differences values of the objective function when objects are temporally removed from the data set.

So, the contents of this paper are the following: in the second section basic concepts and procedures of the heuristic approach to possibilistic clustering are considered in brief, the third section includes describing a novel technique for outlier detection and an illustrative example, in the fourth section preliminary conclusions are discussed and some perspectives for future investigations are discussed.

II. AN OUTLINE FOR A HEURISTIC APPROACH TO POSSIBILISTIC CLUSTERING

The first subsection of the section consists in basic concepts of heuristic approach to possibilistic clustering. Heuristic algorithms of possibilistic clustering are enumerated in the first subsection.

A. Basic Definitions of the Heuristic Approach to Possibilistic Clustering

A heuristic approach to possibilistic clustering was proposed in [12] and the approach was developed in other publications [13], [14], [15], [16]. The essence of the proposed heuristic approach to possibilistic clustering is that the sought clustering structure of the set of observations is formed based directly on the formal definition of fuzzy cluster and possibilistic memberships are determined also directly from the values of the pairwise similarity of observations. A concept of the allotment among fuzzy α-clusters is basic concept of the approach and the allotment among fuzzy α-cluster is a special case of the possibilistic partition which was introduced in [1]. All basic definitions and notations of the heuristic approach to possibilistic clustering can be founded in [12].

Heuristic algorithms of fuzzy clustering display low level of complexity and high level of essential clarity. Heuristic clustering algorithms which are based on a definition of the cluster concept are called algorithms of direct classification or direct clustering algorithms. Let us remind the basic concepts of the heuristic approach to possibilistic clustering.

Let $X = \{x_1,...,x_n\}$ be the initial set of elements and $T : X \times X \rightarrow [0,1]$ some binary fuzzy relation on X with $\mu_T(x_i, x_j) \in [0,1]$, $\forall x_i, x_j \in X$ being its membership function. Fuzzy tolerance is the fuzzy binary intransitive relation which possesses the symmetry property and the reflexivity property.

Let α be the α-level value of the fuzzy tolerance T, $\alpha \in (0,1]$. Columns or rows of the fuzzy tolerance matrix are fuzzy sets $\{A^1,...,A^n\}$ on X. Let A^l, $l \in \{1,...,n\}$ be a fuzzy set on X with $\mu_{A^l}(x_i) \in [0,1]$, $\forall x_i \in X$ being its membership function. The α-level fuzzy set $A^l_{(\alpha)} = \{(x_i, \mu_{A^l}(x_i)) \mid \mu_{A^l}(x_i) \geq \alpha, x_i \in X\}$ is fuzzy α-cluster. So, $A^l_{(\alpha)} \subseteq A^l$, $\alpha \in (0,1]$, $A^l \in \{A^1,...,A^n\}$ and $\mu_{A^l}(x_i)$ is the membership degree of the element $x_i \in X$ for some fuzzy α-cluster $A^l_{(\alpha)}$, $\alpha \in (0,1]$, $l \in \{1,...,n\}$. The membership degree will be denoted μ_{li} in further considerations. Value of α is the tolerance threshold of fuzzy α-cluster elements. The membership degree of the element $x_i \in X$ for some fuzzy α-cluster $A^l_{(\alpha)}$, $\alpha \in (0,1]$, $l \in \{1,...,n\}$ can be defined as a

$$\mu_{li} = \begin{cases} \mu_{A^l}(x_i), & x_i \in A^l_\alpha \\ 0, & otherwise \end{cases}, \tag{1}$$

where the α-level $A^l_\alpha = \{x_i \in X \mid \mu_{A^l}(x_i) \geq \alpha\}$, $\alpha \in (0,1]$ of a fuzzy set A^l is the support of the fuzzy α-cluster $A^l_{(\alpha)}$, $A^l_\alpha = Supp(A^l_{(\alpha)})$. The membership degree defines a possibility distribution function for some fuzzy α-cluster $A^l_{(\alpha)}$, $\alpha \in (0,1]$, and this possibility distribution function is denoted $\pi_l(x_i)$.

Let $\{A^1_{(\alpha)},...,A^n_{(\alpha)}\}$ be the family of fuzzy α-clusters for some $\alpha \in (0,1]$. The point $\tau^l_e \in A^l_\alpha$, for which

$$\tau^l_e = \arg \max_{x_i} \mu_{li}, \ \forall x_i \in A^l_\alpha, \tag{2}$$

is called a typical point of the fuzzy α-cluster $A^l_{(\alpha)}$, $\alpha \in (0,1]$, $l \in [1,n]$. Obviously, a fuzzy α-cluster can have several typical points. That is why symbol e is the index of the typical point.

Let $R^\alpha_z(X) = \{A^l_{(\alpha)} \mid l = \overline{1,c}, 2 \leq c \leq n\}$ be a family of fuzzy α-clusters for some value of tolerance threshold α, which are generated by a fuzzy tolerance T on the initial set of elements $X = \{x_1,...,x_n\}$. If condition

$$\sum_{l=1}^{c} \mu_{li} > 0, \ \forall x_i \in X, \tag{3}$$

is met for all $A^l_{(\alpha)}$, $l = \overline{1,c}$, $c \leq n$, then the family is the allotment of elements of the set

$X = \{x_1, ..., x_n\}$ among fuzzy α-clusters $\{A_{(\alpha)}^l, l = \overline{1, c}, 2 \leq c \leq n\}$ for some value of the tolerance threshold α. It should be noted that several allotments $R_z^\alpha(X)$ can exist for some tolerance threshold α. So, symbol z is the index of an allotment.

Thus, the problem of cluster analysis can be defined as the problem of discovering the unique allotment $R_c^*(X)$, resulting from the classification process and detection of fixed or unknown number c of fuzzy α-clusters can be considered as the aim of classification.

B. Heuristic Algorithms of Possibilistic Clustering: A Brief Review

Direct heuristic algorithms of possibilistic clustering can be divided into two types: relational versus prototype-based. A fuzzy tolerance relation T matrix is a matrix of the initial data for the direct heuristic relational algorithms of possibilistic clustering and a matrix of attributes is a matrix of the initial data for the prototype-based algorithms. In particular, the group of direct relational heuristic algorithms of possibilistic clustering includes

- the D-AFC(c)-algorithm which is based on the construction of an allotment among an a priori given number c of partially separate fuzzy α-clusters [12];
- the D-PAFC-algorithm which is based on the construction of an principal allotment among an unknown minimal number of at least c fully separate fuzzy α-clusters [12];
- the D-AFC-PS(c)-algorithm which is based on the construction of an allotment among an a priori given number c of partially separate fuzzy α-clusters in the presence of labeled object [12];
- the D-AFC(α)-algorithm which is based on the construction of an allotment among an a priori unknown number c of partially separate fuzzy α-clusters with respect to the given minimal value α of tolerance threshold [13];
- the D-AFC(u)-algorithm which is based on the construction of an allotment among an a priori unknown number c of partially separate fuzzy α-clusters with respect to the given maximal number u of elements in every class [14].

Moreover, the FG-AFC-algorithm of heuristic possibilistic clustering based on fuzzy tolerance graph decomposition was proposed in [15].

On the other hand, the family of direct prototype-based heuristic algorithms of possibilistic clustering includes [12]

- the D-AFC-TC-algorithm which is based on the construction of an allotment among an a priori unknown number c of fully separate fuzzy α-clusters;
- the D-PAFC-TC-algorithm which is based on the construction of a principal allotment among an a priori unknown minimal number of at least c fully separate fuzzy α-clusters;
- the D-AFC-TC(α)-algorithm which is based on the construction of an allotment among an a priori unknown number c of fully separate fuzzy α-clusters with respect to the minimal value α of the tolerance threshold.

The hierarchical H-AFC-TC-algorithm which is based on the construction of a hierarchy of allotments among an a priori unknown number c of fully separate fuzzy α-clusters was also proposed in [12].

It should be noted, that these prototype-based heuristic algorithms of possibilistic clustering are based on the transitive closure of the initial fuzzy tolerance. New direct prototype-based heuristic algorithms of possibilistic clustering were proposed in [16] and the family of algorithms is based on the TAGA-algorithm which is calculating different kinds of transitive approximation of the initial fuzzy tolerance [17]. So, the family of prototype-based algorithms includes

- the D-AFC-TAGA-algorithm which is based on the construction of an allotment among

an a priori unknown number c of fully separate fuzzy α -clusters;

- the D-PAFC-TAGA-algorithm which is based on the construction of a principal allotment among an a priori unknown minimal number of at least c fully separate fuzzy α -clusters;
- the D-AFC-TAGA(α)-algorithm which is based on the construction of an allotment among an a priori unknown number c of fully separate fuzzy α -clusters with respect to the minimal value α of the tolerance threshold.

All prototype-based heuristic possibilistic clustering algorithms based on a transitive closure of an initial fuzzy tolerance relation are particular versions of corresponding prototype-based heuristic possibilistic clustering algorithms which based on the calculation of a transitive approximation of a fuzzy tolerance.

III. A NOVEL TECHNIQUE FOR OUTLIERS DETECTION

The proposed technique for outlier detection based on a heuristic algorithm of possibilistic clustering is considered in the first subsection. The second subsection of the section includes a consideration of an illustrative example of application of the proposed technique for outlier detection.

A. A General Plan of the Proposed Technique

Let us consider in detail a plan of the technique for outlier detection which was outlined in [18]. The presented very simple technique is based on sequential application of the D-PAFC-TAGA-algorithm of heuristic possibilistic clustering to a data set.

The technique is based on the assumption, that the cardinality of the outlier class is given a priory by analyst, ϕ . Moreover, the cardinality of a support of each unique outlier is equal one. There is the following seven-step procedure for outlier detecting:

1. The initial data set $X = \{x_1, \ldots, x_n\}$ should be processed by the D-PAFC-TAGA-algorithm by choosing a suitable distance $d(x_i, x_j)$ for fuzzy sets;

2. The support $Supp(A^l_{(\alpha)}) = A^l_\alpha$, $l \in \{1, \ldots, c\}$, $\alpha \in (0,1]$ should be detected for each fuzzy α -cluster in the constructed allotment among fuzzy α -clusters $R^*_c(X)$;

3. The cardinality $card(A^l_\alpha)$ of each support A^l_α , $l \in \{1, \ldots, c\}$ should be calculated;

4. The following condition is checked:
   if the condition $card(A^l_\alpha) < \phi$ is met for some fuzzy α -cluster $A^l_{(\alpha)} \in R^*_c(X)$,
   then the fuzzy α -cluster $A^l_{(\alpha)} \in R^*_c(X)$ is the outlier class and the condition $x_i \in A^l_\alpha$ is also met and these elements can be considered as outliers;

5. The following condition is checked:
   if the condition $card(A^l_\alpha) = 1$ is met for the fuzzy α -cluster $A^l_{(\alpha)} \in R^*_c(X)$,
   then the corresponding object $x_i \in A^l_\alpha$, $i \in \{1, \ldots, n\}$ should be identified as an outlier,
   else go to step 6;

6. The detected small fuzzy α -cluster $A^l_{(\alpha)} \in R^*_c(X)$, or each fuzzy α -cluster $A^l_{(\alpha)}$ from the detected family of small fuzzy α -clusters $\{A^l_{(\alpha)}\} \subset R^*_c(X)$, $card(A^l_\alpha) > 1$ should be processed by the D-PAFC-TAGA-algorithm and go to step 4;

7. The following condition is checked:
   if cardinality of each sub cluster of the outlier class $A^l_{(\alpha)} \in R^*_c(X)$ is equal one,
   then stop,
   else go to step 6.

In other words, if a fuzzy α-cluster includes only its unique typical point then the fuzzy α-cluster can be interpreted as an outlier. The distance for fuzzy sets $d(x_i, x_j)$ and the cardinality of the outlier class ϕ are parameters for the described technique.

The effectiveness of the proposed technique can be illustrated by an example, which is presented in the next subsection.

B. An Illustrative Example

Let us consider the simple illustrative example which was considered by Rehm, Klawonn and Kruse in [9]. The synthetic data set contains two good clusters and some outliers. The data are presented in Table I.

TABLE I
THE SYNTHETIC DATA SET

Numbers of objects, i	$\hat{x}^1$	$\hat{x}^2$	Numbers of objects, i	$\hat{x}^1$	$\hat{x}^2$	Numbers of objects, i	$\hat{x}^1$	$\hat{x}^2$
1	-10.44	-1.33	16	-3.68	-1.73	31	27.98	-4.01
2	14.75	-2.09	17	-8.90	-3.05	32	38.29	-0.04
3	6.78	-1.03	18	2.24	-2.04	33	22.22	-3.63
4	5.25	-0.87	19	2.91	-7.08	34	32.33	-7.45
5	-3.84	1.09	20	4.28	1.14	35	51.01	4.35
6	-1.29	8.42	21	36.04	-1.82	36	37.20	-1.33
7	-6.86	0.60	22	31.40	6.71	37	29.25	-7.83
8	-6.34	3.25	23	38.79	5.04	38	39.43	-1.97
9	-4.47	10.40	24	26.78	1.07	39	33.58	-0.72
10	2.95	-1.70	25	29.72	-1.50	40	40.18	-11.67
11	9.21	3.65	26	33.74	1.28	41	30.00	-18.00
12	6.74	1.47	27	28.51	-0.95	42	-20.00	-1.00
13	-2.46	-4.25	28	41.17	-0.40	43	15.00	-20.00
14	-10.89	-12.67	29	42.47	3.50	44	25.00	-22.00
15	1.19	-11.89	30	36.18	-3.98	45	-5.00	-25.00

This data were pre-processed according to a formula:

$$x_i^{t_1} = \frac{\hat{x}_i^{t_1} - \min_i \hat{x}_i^{t_1}}{\max_i \hat{x}_i^{t_1} - \min_i \hat{x}_i^{t_1}}, \ t_1 = 1, \ldots, m_1, \ i, j = 1, \ldots, m_1 \tag{4}$$

and the normalized Euclidean distance [19] was selected as a parameter for the D-PAFC-TAGA-algorithm:

$$e(x_i, x_j) = \sqrt{\frac{1}{m_1} \sum_{t_1=1}^{m_1} \left(\mu_{x_i}(x^{t_1}) - \mu_{x_j}(x^{t_1}) \right)^2}, \ i, j = 1, \ldots, n. \tag{5}$$

The threshold ϕ was selected in the experiment to n/c, where the number of clusters c is equal 2. The proposed technique was applied to the presented data set and the result is shown in Figure 1.

In the first place, the outlier class $A_{(\alpha)}^2 = \{x_{41}, x_{43}, x_{44}, x_{45}\}$ was detected in the first iteration of the presented technique. This class was processed by the D-PAFC-TAGA-algorithm at next iterations of the proposed technique. At finally, the object x_{45} was identified as an outlier in the first place. The object x_{41} was identified as an outlier in the second place. Objects x_{43} and x_{44} were detected as outliers in the third and fourth place consequently. So, outliers were detected as an ordered sequence $x_{45} \prec x_{41} \prec x_{43} \prec x_{44}$. The objects of "good" clusters are denoted in Figure 1 by symbol ● and detected outliers are denoted in Figure 1 by symbol □.
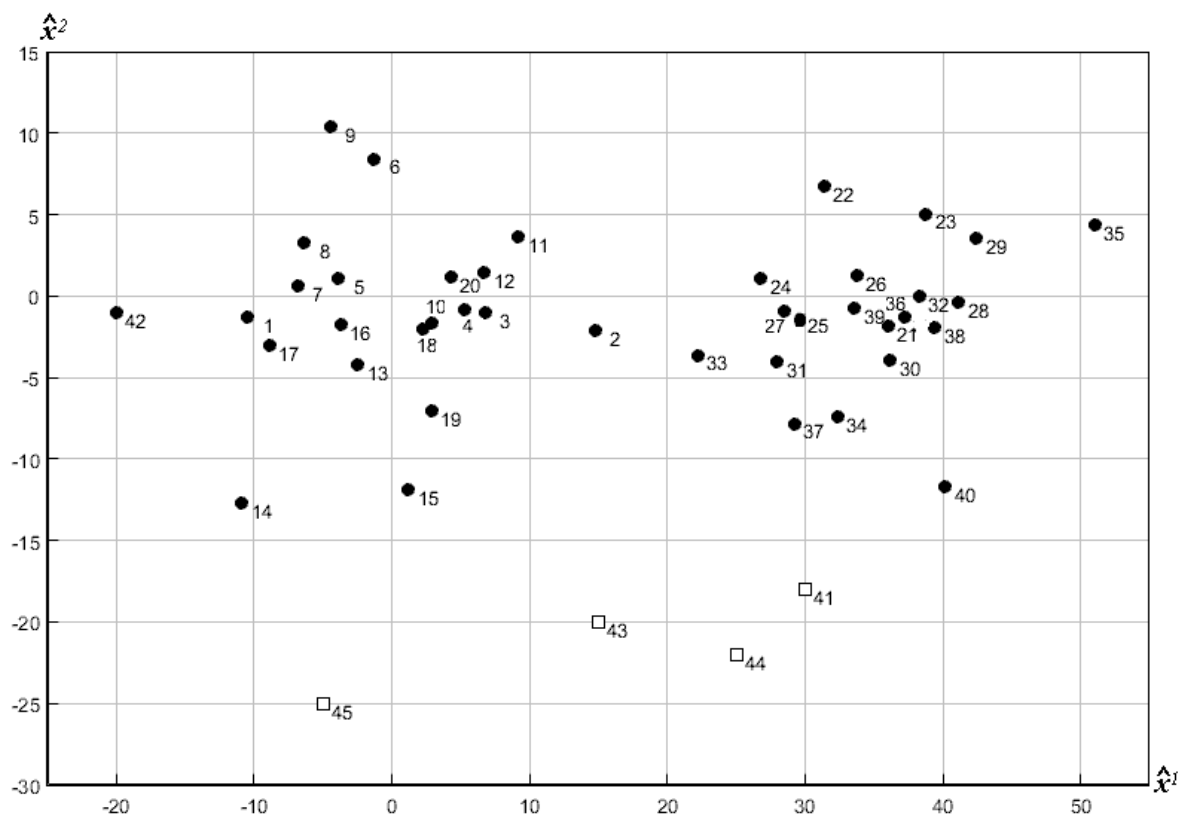
Fig. 1 The data set and detected outliers

IV. Conclusions

Preliminary conclusions are discussed in the first subsection of the section. The second subsection deals with the perspectives on future investigations.

A. Discussions

Outliers can contain information about abnormal behavior of the system and their correct identification is significant. So, the novel technique for outlier detection is proposed in the paper. The technique is based on sequential application of the D-PAFC-TAGA-algorithm to the data set. That is why the ordered sequence of detected outliers is a principal advantage of the proposed technique. This approach can be useful, for example, for assignment of air targets in anti-aircraft defense systems.

The results of application of the proposed technique to the Rehm, Klawonn and Kruse two-dimensional artificial data set [9] show that the technique is an effective tool for solving the outlier detection problem in the framework of exploratory data analysis.

B. Perspectives

Let us consider some perspectives for further investigations. Firstly, the proposed technique for outlier detection should be extended for a case of the large data set. Secondly, the sequential approach to outlier detection based on heuristic possibilistic clustering should be extended for a case of the relational initial data. Thirdly, the idea of sequential outlier detection can be used in the method of outlier detection in the interval-valued data [20].

These perspectives for investigations are of great interest both from the theoretical point of view and from the practical one, as well.

ACKNOWLEDGMENT

REFERENCES

[1] R. Krishnapuram and J.M. Keller, "A possibilistic approach to clustering", *IEEE Transactions on Fuzzy Systems*, vol. 1 (2), pp. 98-110, 1993.

[2] F. Höppner, F. Klawonn, R. Kruse, and T. Runkler, *Fuzzy Cluster Analysis: Methods for Classification, Data Analysis and Image Recognition*. Wiley: Chichester, 1999.

[3] V.J. Hodge and J. Austin, "A survey of outlier detection methodologies", *Artificial Intelligence Review*, vol. 22 (2), pp. 85-126, 2004.

[4] H. Shen, J. Yang, and S. Wang, "Outlier detection by weighed mercer-kernel based fuzzy clustering algorithm", *Iranian Journal of Electrical and Computer Engineering*, vol. 4 (2), pp. 129-136, 2005.

[5] R. Li and M. Mukaidono, "Gaussian clustering method based on maximum-fuzzy-entropy interpretation", *Fuzzy Sets and Systems*, vol. 102 (2), pp. 253-258, 1999.

[6] S.T. Wang, F.L. Chung, Z. Deng, D. Hu, and X. Wu, "Robust maximum entropy clustering algorithm with its labeling for outliers", *Soft Computing*, vol. 10 (7), pp. 555-563, 2006.

[7] V.N. Vapnik, *Statistical Learning Theory*. Wiley: New York, 1998.

[8] J. Łęski, "Robust possibilistic clustering", *Archives of Control Sciences*, vol. 10 (3/4), pp. 141-155, 2000.

[9] F. Rehm, F.Klawonn, and R. Kruse, "A novel approach to noise clustering for outlier detection", *Soft Computing*, Vol. 11 (5) pp. 489-494, 2007.

[10] J.-T. Jeng, C.-C. Chuang, and C.-J. Juan, "Robust interval competitive clustering algorithm with outliers", *International Journal of Fuzzy Systems*, vol. 12 (3), pp. 227-236, 2010.

[11] M. Al-Zoubi, A. Al-Dahoud, and A.A. Yahya, "New outlier detection method based on fuzzy clustering", *WSEAS Transactions on Information Sciences and Applications*, vol. 7 (5), pp. 129-136, 2010.

[12] D.A. Viattchenin, *A Heuristic Approach to Possibilistic Clustering: Algorithms and Applications*. Heidelberg: Springer-Verlag, 2013.

[13] D.A. Viattchenin, "A novel heuristic algorithm of possibilistic clustering for given minimal value of the tolerance threshold", *Journal of Information, Control and Management Systems*, vol. 13 (2), pp. 161-174, 2015.

[14] D.A. Viattchenin, "Heuristic possibilistic clustering for detecting optimal number of elements in fuzzy clusters", *Foundations of Computing and Decision Sciences*, vol. 41 (1), pp. 45-76, 2016.

[15] Viattchenin, D.A., Nikolaenya, E., Damaratski, A. "A fuzzy grapf-based heuristic algorithm of possibilistic clustering", *Communications on Applied Electronics*, 3 (7), (2015), pp. 13-23.

[16] D.A. Viattchenin and A. Damaratski, "Direct heuristic algorithms of possibilistic clustering based on transitive approximation of fuzzy tolerance", *Informatica Economicá*, vol. 17 (3), pp. 5-15, 2013.

[17] P. Dawyndt, H. De Meyer, and B. De Baets, "UPGMA clustering revisited: a weight-driven approach to transitive approximation", *International Journal of Approximate Reasoning*, vol. 42 (3), pp. 174-191, 2006.

[18] D.A. Viattchenin, "Sequential outliers detection by using heuristic possibilistic clustering", *Proceedings of the XI International Conference on Computer Data Analysis and Modeling (CDAM'2016)*, Minsk, Belarus, pp. 317-318, September 2016.

[19] A. Kaufmann, *Introduction to the Theory of Fuzzy Subsets*. New York: Academic Press, 1975.

[20] D.A. Viattchenin, "Detecting outliers in interval-valued data using heuristic possibilistic clustering", *Journal of Computer Science and Control Systems*, vol. 5 (2), pp. 39-44, 2012.

Approaches to the Integration of the System of Education of Belarus into the World Educational Space

U.A. Beizerau

Abstract—The research theoretically grounds the development of the processes of integration in the world educational space at the end of the XX c. - the beginning of the XXI c. and the model of integration of Belarussian educational system into the world educational space. The essence and the contents of the integration processes in the world educational space are shown; trends of the development of integration processes in the world educational space are discovered; model of integration of Belarussian system of education into the world educational space was created; the peculiarities of integration of Belarussian system of education into the world educational space were defined.

Keywords— integration; the system of education; educational space; globalization; processes of integration

I. INTRODUCTION

Obtaining of independence, political, economic and sociocultural changes in the Republic of Belarus caused need of carrying out transformations in all spheres of activity of society, including educational system, its integration into the World educational space.

The World educational space is a set of all educational institutions, scientific and pedagogical centers, governmental and public organizations for education in various countries, geopolitical regions and on a global scale, their interference and interaction in the conditions of intensive internationalization in different spheres of public life of the modern world.

The Republic of Belarus, being the full member of the world community, one of the UN countries foundresses, the member of a set of international organizations, being in the geographical center of Europe, can't remain away from global processes of integration, including in world educational space.

II. INTEGRATION WITHIN CIS MEMBER STATES

Being the member of the Commonwealth of Independent States, Republic of Belarus has to build an educational system taking into account the integration processes happening in the countries of the former Soviet Union. Heads of the governments of CIS signed on May 15, 1992 the Cooperation agreement in the field of education for carrying out the coordinated educational policy, development of ways and mechanisms of practical implementation of joint multilateral projects. The agreement of Ministers of Education of the CIS countries (October 3, 1992) and creation of constantly operating advisory body – Conferences of Ministers of Education of the State Parties of the CIS, and also laws on formation of the CIS countries, etc. are initial standard and legal base to create scientific and organizational basis for development of integration processes in education. Concluded in April, 1996. An agreement between the Republics of Belarus, Kazakhstan, Kyrgyzstan and the Russian Federation on deepening of integration in economic and humanitarian areas became an important stage on the way of formation of uniform educational space. The instrument for development and system realization of state policy on support of integration processes in the field of education of the CIS countries was

U.A. Beizerau F. Scorina Gomel State University, Gomel, Belarus (e-mail: bejzerov@gsu.by)

approved by authorized bodies of the CIS, providing its legal, organizational, financial, material, personnel, scientific and methodical and information support.

However integration processes in the CIS is only a part of universal processes of globalization and integration. It is possible to note that integration in the CIS is a reintegration, or restoration of that already existed.

With the collapse of the USSR and finding independence by all former soviet republics, their educational systems, continuing to keep bases of the Soviet model, underwent basic changes. Borrowing new of foreign models, one of the best in the world, Soviet educational model was almost completely rejected. This circumstance distanced the former soviet republics from each other more and more.

Existence of the national education systems which developed in the middle of the XX century separately from each other wasn't able to meet requirements which are imposed now by the society. Integration of national education systems into the World educational space with their corresponding unification and updating in connection with the requirement of time became really necessary.

Problems of integration of national education systems are subject of scientific research of pedagogues, philosophers, psychologists, economists, lawyers and sociologists. The tradition of illumination of this problem in the world historiography starts from the second half of the XX century.

Within the center of comparative pedagogics (Moscow) in 1992 the group of researchers studying the development of education in the CIS countries was created.

In 1993 the group prepared and published the analytical report "Legislative, standard and legal base of the development of education in neighboring countries" in which legislative and legal documents on formation of the CIS countries were considered, the analysis of general provisions of legislative and standard and legal base of educational institutions, developments of links of education, social guarantees of realization of the rights of citizens for education, questions of differentiation of competences of the field of education, aspects of management, economy and financing of educational institutions, the international cooperation and foreign policy activity contained in the field of education. The scientific field of research was made by problems of justification of initial positions of steps and levels of continuous education in the CIS countries.

TABLE I
CHANGING STRUCTURE OF BELARUSSIAN HIGHER EDUCATION

Level	Old system	New system
1	Specialist diploma	Bachelor degree
	5 years (6 years – medical specialties) – for full time students, 6 years – for part-time	4 years (6 years – medical specialties) – for full time students, 5 years – for part-time.
2	-----	Master degree
		1-2 years (depending on specialty and form of study)
3	Candidate of Sciences	Candidate of Sciences (PhD)
	3 years – full time, 4 years – part-time	3 years – full time, 4 years – part-time

It is necessary to mention, that processes of integration into the World educational space passed in the CIS countries unevenly, not in the same timeframes, through different stages. Most of the post-soviet countries (including CIS countries) joined the Bolognia declaration in the early days of its existence. The pioneers were Baltic states, Ukraine, Georgia and Russia. For many years (until 2014) Belarus remained the only European state that did not join the common integration processes. Nevertheless, the road map was designed and several steps according it have been implemented (for example transition to the 3 circle structure of higher education – returning undergraduate (bachelor) degrees, changing master-level degrees, turning to the competence-based approach in education, etc.). The changing structure of the Belarussian higher education can be illustrated in the table I.

III. THEORETICAL GROUNDS OF INTEGRATION

At the same time the insufficient theoretical readiness of the sociocultural bases of integration of the Belarusian education system into the World educational space, unilateral emphasis in domestic pedagogical science only on positive or negative sides, insufficiency of complete idea of the realized and unrealized potential of the national educational system for its modern transformation in innovative, open and effective system which will be a harmonious part of European and World educational space became a basis of definition of a subject of our research.

Integration processes take place in education under the influence of the world process of globalization and is the part of universal processes of integration in all spheres. The content of integration processes is directed to unification of educational systems by the principle "Unity in variety" which assumes a community at preservation of national lines, opens their essence. Specifics of integration processes in education are defined by the principles: humanity (reflects a freedom of choice of each person on self-determination in questions of a choice of ways of integration), democracy (assumes equal rights and opportunities in education regardless of race, nation, social origin), mobility (reflects variety of means, ways, integration forms, its flexibility, readiness for fast reorganization of national educational systems), advancing (demands anticipation of possible options and results), openness (assumes expansion of education, attraction to the educational sphere of new experts, extension of the list of specialties and specializations), continuity (designates need of continuous course of integration processes), phasing (assumes existence of the integration stages defined and coordinated at all levels), systemacity (designates that integration processes exists as a system), equalities (provides to each participating party of integration process equal conditions), unity in variety (assumes implementation of integration processes taking into account national and regional features of the countries).

In the conditions of intensive development of science and transition to information society strengthening of attention to the tendencies causing processes of unification of components of educational systems is required. The main tendencies of development of integration processes in world educational space are: growth of scales of education; diversification of education in institutional forms, levels and the contents; cooperation of the states in line with convergence of policy of the governments in the field of education; partnership at all levels of management, including education; internationalization in education; competition in education; versatility, globality, flexibility of education; availability of education; remote education; technologization and computerization of infrastructure of education; through life education; creation of target supranational programs; transformation of traditional systems.

The structural and functional model of integration of the Belarusian education system into the World educational space assumes, on the one hand, integration of educational system of the Republic of Belarus into the World educational space on the basis of the European principles, on the other – implementation of integration within the Union State of Russia and Belarus, "Big

Five" and the CIS; it is realized taking into account the principles of humanity, democracy, mobility, an advancing, openness, a continuity, phasing, systemacity, equality, unity in variety. Integration of the Belarusian education system into the World educational space is carried out step by step: organizational stage assumes adoption of the relevant decisions, signing of international treaties that conducts to integration of education systems; the functional stage includes drawing up the project plan of integration, implementation of necessary reforms of education systems for complete unification with educational systems of other states; the control stage is directed on implementation of the mutual controlling actions over carrying out educational reforms; the correctional stage includes comprehensive monitoring of the carried-out changes, correction, studying of the changes which happened in an education system after carrying out reforms. Functioning of model of integration of the Belarusian education system with educational systems of other states is carried out on the basis of unification of legislative, sociocultural, psychological outlook, structural, material and educational components [3].

Features of integration processes of the Belarusian education system in the World educational space, which allows to define aspects of reforming of the educational system of the Republic of Belarus in the direction of its rapprochement with educational systems of other states: geopolitical (position in the center of Europe; existence of strong communications with the European Union, with Ukraine; membership in the Union of Belarus and Russia; strengthening of processes of globalization in the World; isolation of the educational system of Belarus from the World educational space for more than seventy years in the XX century, etc.); sociocultural (European character of social and cultural spheres; existence of one of the most effective educational systems in the World; powerful scientific and welfare potential; reforming of the education system taking into account features of integration processes, etc.); economic (essential economic potential; transition of educational institutions to market forms of organization and commercialization of educational services, etc.); typically educational (standard legal support; principles, priorities of development of the national education system; functioning of educational institutions of various types and forms of ownership; structure and duration of education at all stages is approached to the duration in the European countries; introduction of two-level system of training of specialists with the higher education; introduction of system of test units, etc.) [3].

Need of reforming of the contents and methodology of existing educational systems became really acute. The idea about discrepancy of the operating education system more often called traditional to modern public requirements, a level of development of production, economy and society in general is put forward. In this regard, changes in educational systems have to be focused on the integration processes happening in world educational space.

According to definition, integration (from Latin integration – restoration, completion, from "integer" – whole) – association in a single whole, streamlining, structuring earlier disconnected, disorder phenomena, parts of the whole. In the course of integration the volume, frequency, intensity of interactions between elements of a system increases, bigger degree of its integrity, stability, an autonomy and efficiency of action are reached. The result of the process of association and unity, condition of harmonious steadiness, ordered functioning of parts (close concepts can be understood as integration: system, structure, organization, and model). Depending on nature of system object integration can be considered at the level of the personality, group, region, city, economic and cultural complex, social system, the commonwealth of the states and the world community. According to the nature of the interacting elements we can distinguish integration groups – social, cultural, etc. [9].

In the context of our research first of all social and cultural integration, shown in interaction of educational systems and spaces as education is a basis of a welfare paradigm are important.

Cultural integration – process of deepening of cultural interactions and interference between the states, national and cultural groups and regions. It can be considered as: 1) process of practical and information exchange between cultural establishments, the centers, creators and

consumers of culture; 2) achievement of bigger compliance and coherence between the various cultural establishments, traditions, national forms of culture, between cultural heritage of society and new achievements of culture; 3) process of the statement of uniform system of values as a live, friendly exchange between cultures at the interstate and interstate levels. Unlike social integration, integration cultural doesn't assume "alignment" of norms and values of culture, achievement of their uniformity. In the conditions of modern democratic cultural integration is aimed at the most free development of each of cultures by means of all others. Cultural integration is complicated by distinctions of outlooks, religions, the unconscious cultural installations developed during the millennia. The management of processes of cultural integration assumes tolerance and careful attitude to different views and customs, aspiration to understand each other; use of various forms of cultural cooperation, wide circulation of international modes of production, communications, trade, transport, scientific and technical cooperation; introduction of uniform programs of international education and education taking into account national features of culture [9].

Social integration – process of establishment of optimum communications between social institutes, groups, echelons of power and management; development of spatial and territorial system of communications between various regions, development of uniform ideology by society. Public work of people within this or that way of production which defines its ways, mechanisms, forms basis of social integration. In antagonistic society at the totalitarian modes social integration happens most often by infringement of the rights and freedoms of the social groups, ethnic minorities, liberation and revolutionary movements, is carried out by violence and terror. In democratic society social integration is based on free, equal and mutually beneficial cooperation of economic regions and states. It assumes effective division of labor, exchange of scientific and technical experience and information, coordination of plans of social development, establishment of uniform system of standards in all branches, different creation of the advisory and coordination centers. The purpose of social integration – alignment of indicators of economic and social development of the countries and regions.

The available publications testify that now the problem of integration in education became independent object of research in pedagogical science and practice.

Special relevance of a problem of integration in education was gained in the 1990th. A distinctive feature of researches of this period are the works giving a complete picture of integration at rich historiographical material, based on use of system and historical methods. Undoubted advantage of these works is the introduction to science a large number of historiographic sources, statistical materials, deep scientific validity of the received results. At the same time the problem of integration was considered in them against the general review of a condition of education.   From the middle of the 1990th the problem of integration becomes independent object of the scientific analysis.

IV. ANALYSES OF THE PROBLEM

Publications of the famous researcher in the field of comparative pedagogics B. L. Vulfson consecrate problems of the World educational space. The author opens communication of the integration processes proceeding in education systems of the countries of the world with globalization processes, characteristic for all spheres of activity of society; points to positive and negative aspects of the studied phenomenon. In modern World educational space demands orientation to development tendencies of education in various countries. The researcher pays main attention to the quantitative indices, characterizing modern World educational space, provides the development analysis of the educational sphere in Russia, carries out comparison of educational systems of other countries, analyzes demographic prerequisites of the changes happening in the World educational space. The World educational space is represented as a uniform formation which all national education systems will enter gradually [11].

In the research conducted by V. V. Anisimova and O. G. Grokholsky problems of creation of uniform educational space in the context of integration of Russia and Belarus are considered and methodological bases of integration of education systems and principles of integration are development. They developed methodological installations of integration of educational systems of Russia and Belarus which are applicable (with insignificant correction) and in integration of education systems of all countries of the World community into educational space.

Many scientists are investigating the problem of transformation of the Belarusian educational standards to the norms of the European Union, note that European Higher education institutions faced the same problems, and transition to new quality of education according to the requirements of the XXI century is so difficult for them, as for HEIs of Belarus. Educational systems have to develop on the basis of cooperation taking into account the experience which exists in different educational systems [7].

Researchers L.S. Onokoy and V. V. Senashenko study the problems and prospects of integration of the higher school of Russian Federation into European education system. According to the researches, Russian labor market was not ready to employ bachelors and masters. According to some experts, the two-level system of higher education accepted at the level of legal acts was unclaimed by economic and governmental institutions of the country. Moreover, "it was sawn-off considerable part of educational community". Authors place emphasis on the point that education is based on cultural and pedagogical national traditions and priorities, has deep historical roots, is formed taking into account mentality of the people living in the country. Any innovations in education have to be carried out after careful methodical, technological, marketing study in the presence of standard support and only in that case when society is ready to accept them. High social importance of education doesn't allow to turn it into a platform for rash experiments [7].

A.N. Dzhurinsky's works raise questions of internationalization of higher education in the modern World. The researcher considers that internationalization of the higher education is an objective consequence of globalization of the modern World and at the same time large resource of acceleration of integration of public and economic life, elimination of national isolation; points to that fact that process of internationalization was difficult and contradictory. Analyzing the course of integration processes in some countries of the West, A.N. Dzhurinsky draws parallels with the processes proceeding in the Russian higher education, describes and analyzes the projects existing in the European Union directed on rapprochement of education systems. The important advantage of publications of the scientist is the analysis of problems and prospects of integration of educational systems [5].

In the research of A.P. Liferov, problem of reintegration of educational space of the CIS countries is traced. The author notes that in the process of integration CIS gains lines of a multiplelness; traces features of course of integration processes in border regions; defines the factors promoting expansion of interstate cooperation in the field of education, analyzes problems and prospects of development of integration processes in each of the countries of "border-zone". The scientist considers that in "reintegration of educational space of the CIS the Western vector dominates, i.e. at reunion of education systems of the former republics of the Soviet Union, main "beacon" are educational systems of the countries of the West, generally EU countries" [6].

In scientific articles and monographs devoted to integration of national educational systems special stress is given to systems' transformation. Globalization, according to the western analysts, is the main factor of integration of educational systems into educational space and acts as one of the most important parameters of existence of human society. It is explained by rapid growth of the population of Earth, and also uneven distribution of income. During the period between 1960 and 1994 the number of people living below the poverty line increased in Europe, Central Asia, Latin America, Caribbean countries and the countries of Africa [8].

The general expenses for education comprise 80 billion dollars a year now. The World community through joint efforts already achieved some progress in the educational perspective. So, as an example, number of children attending school increased from 599 million in 1990 to 681 million in 1998. Since 1990 every next year the number of the children attending school increases by 10 million. East Asia, the Pacific region and Caribbean countries are close to introduction of universal primary education. The number of children who aren't attending school for the reasons stated above around the World decreased from 127 million in 1990 to 100 million in 2004 and continue to decrease. Nevertheless, in the countries of Africa to the South from Sahara, the number of the children who aren't attending school grew up on 17% in connection with the sharp growth of population. In Latin America and Caribbean countries, for example, the number of children who aren't covered by primary education decreased for the specified period twice – with 11,4 to 4,8 million. The number of children visiting preschool institutions increased for the last decade by 5%. The number of competent adults doubled from 1970 to 1998 and increased from 1, 5 billion to 3,3 billion. Today 85% of men and 74% of women around the World are able to write and read. 87% of youth (from 15 to 24 years) are around the World competent. However, despite notable progress in education, 17% of World's population or 785 million adults remain illiterate, from which – 63,8% – women [12].

Since 1990 to 2005 the number of illiterate decreased (from 875 million). Thus, positive role of globalization on development and improvement of national educational systems is shown. Globalization also has positive impact on integration of education systems of the countries of the World into uniform world educational space. This process will lead to emergence of the general educational standards, uniform curricula and programs, uniform system of estimation of knowledge and abilities of students, deeper and closer interaction of educational institutions [12].

Processes of modernization of various spheres of public life of the Republic of Belarus at the end of XX – the beginning of the XXI centuries are caused by internal processes of social and economic character, and global tendencies connected with transition to an information and technogenic phase of development of human civilization. At the same time it should be noted that these processes have contradictory nature which is caused by factors of political and ideological character. Contradictions between ethnocultural processes in the modern former Soviet Union and integration processes in culture amplify due to so-called paradox of ethnicity, so characteristic for modern World, – when in the process of deepening of internationalization of culture, the level of national consciousness increases.

Educational situation of Post-Soviet Belarus was a consequence as socio-political differentiation within uniform allied space, and the accruing process of globalization of public life on a global scale. Caused by transition to a civilization of technogenic and information type, process of globalization was reflected in all spheres of activity, including in dynamics of development of educational system. At the same time it should be noted that the idea of a globalism by the internal nature is peculiar to domestic philosophical and pedagogical thought [4].

At the present stage there is a consolidation of welfare identity of Belarus within the global tendencies. Initial general designation of various initiatives, actions, organizations, movements which purpose is achievement of visible unity of people. Now it is rather an idea, search, aspiration to find unity), i.e. aspiring to a unification on the principles of moral.

Antipode of integration processes is the processes of nationalization, or geoethnoregional fragmentation when the majority of the countries of the World have their national traditions including educational which also develops eventually. These processes happened and happen also in Belarus.

In the conditions of the World integration and prompt development of global information technologies, the universal communicative environment is erasing borders between continents,

states, and people. The main objective of education is connected with development, varieties of information streams, integration of science and culture. The state policy in the field of education has to be implemented in the interests of formation of harmoniously developed, socially active, creative person and as one of factors of economic and social progress of society on the basis of the priority of education proclaimed in the Republic of Belarus [10].

Modern process of the state decentralization and regionalization causes need of search for the educational sphere representing the defining side of culture of society, strategy of reforming and the concrete directions of an exit from the created crisis. Because the educational paradigm changed, it is necessary to develop innovative education which essence can be found in new pedagogics, new educational processes, new technologies. Such transformation of education system which would promote updating of all spheres of life of society is necessary. Feature of modern social and economic reforming in Belarus is the mobility and flexibility demanding some theoretical judgment of the processes [2].

Research works of many scientists are devoted to the problems of functioning of educational system of modern Belarus. Authors created scientific and pedagogical concepts within the development strategy of educational systems are defined, mechanisms of their functioning in the innovative mode taking into account fundamental tendencies are offered. It should be noted that, despite of considerable differences in a big variety of approaches, most of the researchers incline to that any national educational system has to be a component of uniform educational space of the region (continent, the World).

Main objectives and problems of education in modern Belarus are seen in ensuring historical continuity of generations, preservation, distribution and development of national culture; education of patriots, the citizens of the constitutional, democratic, social state respecting the rights and personal freedoms and morality; formation of complex outlook and modern scientific outlook, cultural development of the interethnic relations. At the same time, growing threat of violation of unity of educational space regarding training in a number of subjects, as the main problem the is noted. Implementation of legislation in the field of education means, mainly, updating of its contents and introduction of educational standards. Certain difficulties in the solution of the matter are dictated by that subjects of the world community realize a national and regional component of the content of education in any forms that complicates reduction of plans and programs for training courses to the uniform standard and detection of efficiency of the educational process.

The principle of regionalization acts as the dominating factor of development of educational systems which realization designates a new stage of formation of variable education in the modern World now. This principle allows ethnical and regional educational systems to act as subjects of strategy of development according to geo-economic and sociocultural features inherent in the region. As E.A. Sobolev and V. I. Spirin specify, the holistic approach allowing to establish organic or functional connection between whole and the parts making it, to reveal contradictory tendencies of regionalization and integration of education, its development and self-development, variability and integrity of educational space forms the methodological basis of realization of the principle of regionalization. The region acts as a steady social and geographical concept which gets more concrete sense in modern educational space. At the same time, there is a question concerning how legally reduce national and regional component of the content of education to the needs of a certain country and people in the sphere of reproduction of sociocultural values [2].

The main contradiction in development of ethno-regional educational systems of the modern World seems in discrepancy of ethno-regional and poly-cultural factors of creation of the content of education. On the one hand, development of educational system is urged to keep national identity, on another – to create conditions for "injection" into the World cultural and educational space.

We consider several steps should be taken for creation of uniform cultural and educational space: 1.To create the uniform concept of the content of education. 2. At the international level to raise a question of creation of textbooks and manuals, to develop flexible and variable mechanisms of preparation of pedagogical shots in the system of secondary education taking into account the global tendencies in the cultural and educational sphere. 3.   "Global regionalization" of public life caused need of reconsideration of a role, situation and prospects of further development of domestic education. As for an educational situation of modern Belarus strengthening of an ethnization of the content of education, increase of a role of national culture in the course of socialization and an interiorization of the personality is characteristic, ethnocultural factor has to act as a basis for polycultural education. The leading principle of realization national in polycultural education – the sociocultural context of development of the personality assuming the maximum accounting of concrete national realities and definition of their place and the importance in the international world. On the basis of this purpose the educational strategy directed on formation of the personality capable to active and effective activity in the multinational and polycultural environment having the developed feeling of understanding and respect of various national cultures is defined. This educational strategy defines the following educational tasks: deep and comprehensive mastering bases of National culture that is an indispensable condition of integration into other cultures; formation of ideas of variety of national cultures, education of the positive relation to the cultural distinctions providing conditions for self-realization of the personality; familiarizing with bases of world culture, disclosure of the objective reasons of process of globalization, interdependence of the people in modern conditions. Modern polycultural education has to be based on the basis of the accounting of the following principles): the principle of a dialectic inclusiveness of national culture in system world, expressing system of common cultural imperatives which are inherent equally in all ethnocultural educations; the principle of a historical and cultural and civilization orientation of the national education assuming need of disclosure of historical conditionality of the phenomena of the past and the present, studying of folklore, national art, customs and traditions. Thus the concept "national culture" gets extensive, comprehensive and multidimensional sense as the integrative category uniting various aspects of identification of society in domestic and world culture. As mentality is formed as interaction of the national philosophies, national religion and folklore, these components act as reference points in creation of pedagogical theories of both national, and political cultural education; the principle of polycultural identification and self-updating of the personality based on inclusion in the regional content of formation of knowledge of the person and society; the principle of globality of the cultural and educational process which is responsible for development of complete polycultural outlook, communicative features of the personality in the modern world; the principle of polycultural tolerance and inter educational prospect reflecting the mechanism of ethnocultural identification of the personality to structure of harmonization of the interethnic relations.

Modern polycultural education has not only to be guided by the comparative analysis of various ethnocultural modes, but in every possible way use diachronic approach allowing to track stages of formation of system of spiritual and moral values, to see in historical and retrospective aspect the invariant and specific phenomena in development of a certain type of culture, to feel "spiritual style of an era" and to reveal artifacts in development of this ethnoculture. Diachronic approach in structure of polycultural education is directed on identification and the analysis of the "cultural stereotypes" representing set of the most characteristic features of culture. As the cultural stereotype represents set interacting with each other and complementary elements, consideration of each of them inevitably brings to the level of studying of the system. This approach answers the idea of humanitarization of education as it is directed on complete studying of culture [2].

Organized thus polycultural education relies on interdisciplinary creative approach and is directed on formation of a culturological reflection as a result of which the learning subject rises over various ethnocultural paradigms of vision of the world and the person. Formation of the polycultural thinking acting as a reflection of modern global tendencies in the educational sphere is the result of this educational process.

V. CONCLUSION

The article showed main tendencies in the development of educational system of the Republic of Belarus, analyzed the processes happening in education and global trends. The topic developed in the article needs further investigation with the new events and challenges which take place in the World educational space, brought by globalization.

Integration processes take place in education under the influence of the world process of globalization and is the part of universal processes of integration in all spheres. For many years Soviet Union remained isolated from the World educational community. Until the first decade of the XXI century Belarus remained the only European state, that did not join the common integration processes. Due to the road map, developed by Belarussian side, the process of integration took its real shape. But still there are too many obstacles on the way to complete integration (different terms of secondary education, different approaches to teaching and educational process, contradictions in the vision of formal side of subjects, etc.) which make the process longer.

REFERENCES

[1] A. Belogurov, Influence of process of globalization on formation of regional educational space. Bulletin VIU (2(8), 2002, pp.141-146.
[2] A. Belogurov, Strategy and methodology of sociocultural modernization of regional education. Pedagogics (2), 2012, pp.13-22.
[3] M. A. Biryukova, Globalization: integration and differentiation of cultures. Philosophical sciences (1), 2001, pp. 33-42.
[4] S. I. Buldakov, Philosophy and methodology of education: scientific publication. Kostroma: KGU of N. A. Nekrasov, 2014.
[5] A. N. Dzhurinsky, Internalization of the higher education in the modern world. Pedagogics (3), 2004, pp.83-92.
[6] A.P. Liferov, Education in strategy of multinational corporations. Pedagogics (2), 2005, pp.79-90.
[7] L.S. Onokoy, Russia on the way of integration into the all-European education system: sociology and education. Sociological research (2), 2004, pp. 80-85.
[8] L. Portnoi, Higher Education, Policy and the Global Competition Phenomenon. NY: McMillan, 2014, 266p.
[9] J.Rigg, Is Globalisation Good? Geography Review (4), 2001, pp.36-37.
[10] The big encyclopedic dictionary (philosophy, sociology, religion, an ezotermizm, political economy), Under S.Y. Solodovnikov's edition. MFTsP, 2002, p. 214.
[11] The Education law of the Republic of Belarus (1202-XII): National register of legal acts of the Republic of Belarus (92), 2006, 2/1220.
[12] The SAGE Handbook of International Higher Education. London: Sage, 2015, 536 p.
[13] B. L. Vulfson, World educational space at a boundary of the XXI century. Pedagogics (10), 2002, pp. 3-14.
[14] World education report. To knowledge society. NY: UNESKO publishing, 2015.

Application of GIS-Technologies in Ecological Education

A.S. Sokolov, O.V. Kovaleva

Abstract –Full-time training and distant courses for teachers` continuing education of ecology-related subject corresponding to principle directions of state programmes in the sphere of ecological education within the context of multilevel system of education were developed and implemented by academicians of Educational Institution «Francisk Skorina Gomel State University» jointly with partners (Educational Institution «Vitebsk State Technical University» and Educational Institution «Polessky State University») within the framework of EcoBRU project. The article presents the opportunities of the developed course «Application of GIS-technologies in ecological education».

Keywords – ecological education, GIS-technologies.

I. INTRODUCTION

Ecological education and upbringing is a relevant social, economic, psychological pedagogical, humanitarian and ethic problem the essence of which is revealed through the development of environmentally friendly behavior of an individual. There is no need to prove the significance of ecological upbringing. There are two reasons for such particular attention to this subject: a necessity to consider ecological upbringing and education as a continuous systematic process during the whole period of study and relevance of formation of elementary ecological culture. One of the most significant ethical and ecological peculiarities acquired in the process of ecological education and study is ecological responsibility revealed through appropriate, that is, environmentally friendly behavior of a person. Ecology related perceptions enter all the spheres of scientific and social life, serving as a factor which unites the world community during the process of globalization. Ecology becomes a leader of scientific picture of world whereas ecological education is a top-priority and system developing factor for education at large [1].

It is also difficult to overestimate the significance of information and society IT development processes in XXI century. Information is becoming a resource of continuously increasing importance aiming at society development. Development of computer technologies provides both enhancement of already used data processing and constant application of new data, computerization of new areas of knowledge and management. Programme and technical means which implement information technologies in practice are used for information resources access. Great deal of the real world information and practical activities – techniques, economics, ecology, politics, management – are of spatial reference. Spatial information processing systems are of great importance nowadays with the leading role of geographic information systems taken into consideration.

Educational Institution «Francisk Skorina Gomel State University» is engaged in implementation of «Ecological education for Belarus, Russia and Ukraine» (EcoBRU) project within the framework of Tempus programme of TACIS European Community. The project, aiming at qualification enhancement of secondary school, college and university teachers, has as its goal the development of ecological education oriented study programmes in the system of continuous education for teachers and academicians in the context of multilevel system of

A.S. Sokolov, Educational Institution, Francisk Skorina Gomel State University, Gomel, Belarus.
O.V. Kovaleva Educational Institution, Francisk Skorina Gomel State University, Gomel, Belarus, (e-mail: sanakovaleva@mail.ru).

education in Belarus, Russia, Ukraine; testing, implementation, official recognition and gradual use of distant study courses in ecology in the context of multilevel system of education; modernization of the existing study programmes in the context of ecological education and environmental protection.

At present there are scarcely any methodological materials for teachers concerning GIS technologies and its application in ecological education and ecological researches in Belarus. The absolute majority of teachers are unaware of this subject, as well as of GIS-systems methods and the possibilities which they open for study and research activities. Meanwhile the possibilities of ecological mapping, acquisition, visualization and analyzing of various ecological information by means of GIS-technologies are numerous and increasing annually.

Application of GIS-technologies enables its applicants to diversify forms of activities and possibilities in ecological education, organize ecological education and researches in a way corresponding to modern requirements of society and integrated with global trends in this area.

II. CONTENTS OF GIS-TECHNOLOGIES APPLICATION IN ECOLOGICAL RESEARCHES STUDY PROGRAMME

Study of GIS-technologies application in ecological education aims at development of the following skills:

- possibility of independent creation of three-dimension models of any geographic objects, from global to local; having both traditional map and its three-dimension model applicable, learners will be able to master the technique of «catching» spatial three-dimension information on flat two-dimension maps considerably quicker than ever;
- possibility to transform the existing flat two-dimension images (of a map, space image) into three-dimension model promptly;
- thematic mapping using free software and all types of cartographic images, choice and performing of any mapping;
- free acquisition of administrative division maps of any level of all countries;
- acquisition of space images of any territory and of different resolution from various satellites and processed material of space survey from profile organizations resources (for example, relevant cloud map, forest fire map, ice map, map of sea and ocean pollution, Arctic ices dynamics, desert advancing map, etc.);
- possibility of free acquisition and processing of basic materials of space survey (images separated by shooting channels), mastering of many study and research possibilities of processing such materials;
- mastering of dimensional analysis methods realized with GIS-technologies application (calculation of morphometric figures, calculation of distance between objects, buffer zone development, maps layering and resulting maps development, attachment of additional materials – images, text files, video files, audio files, etc. – to e-maps);
- use of geological and social servers, media servers aiming at acquisition of relevant and visual materials on various topics illustrating natural, social and economic peculiarities of different Earth regions;
- mastering of global e-maps activities methods (including animation maps) appearing in the latest years (meteorology and hydrology figures maps, geology maps, forestry maps, political development bearing maps, etc.);
- use of other free sources of ecological and geographical information.

The programme of mastering the subjects dedicated to use in ecological education and researches space survey related materials, as well as the correspondent acquired possibilities, is presented below as an example.

III. ECOLOGICAL RESEARCHES AND PROJECTS IMPLEMENTATION BY MEANS OF EARTH DISTANT SOUNDING MATERIALS

Earth distant sounding materials (space images, digital surface model, etc.) can be used for acquiring information about various territories land management structure, zones of the most transformed areas, principles of ecosystems dimensional dissemination, natural, social and economical objects and phenomena. Mainly these data serve as a foundation for territories ecological evaluation, its dynamics and implementation of study and research projects in learning process.

Learning of space materials application in ecological researches was organized in numerous stages. The first stage is acquisition of space materials from free sources. For this purpose such web-sites as the USA Geological Services (http://earthexplorer.usgs.gov), University of Maryland (http://glcf.umd.edu), Roscosmos geoportal (http://gptl.ru), etc. Multi zone space survey can be acquired through these servers - scenes Landsat 5 TM, Landsat 7 ETM+, Landsat 8 OLI, Aster, EO-1 ALI, EO-1 HYPERION. Kanopus-B, Resurs-DK1, SPOT 5, Formosat-2, Ikonos.

The images represent sets of pictures made in various spectral ranges. For example, scene Landsat 7 comprises panchromatic image with values varying from 520 to 900 nm with 15 m in GSD distance , infrared image with values varying from 10400 to 12500 nm with 60 m in GSD distance and multi spectral image comprising 6 pictures (values: 1-450-515; 2-525-605; 3-630-690; 4 – 760-900; 5 – 1550-1750; 6 – 2080-2350 nm) with 30 m in GSD distance and radiometric resolution corresponding to 8 bit.

The second stage is synthesis of multispectral image colour picture and its further thorough analysis. We propose to consider MultiSpec programme (https://engineering.purdue.edu/~biehl/MultiSpec/) as a software for such purposes. The programme is designed for computer processing of multispectral and hyperspectral images and has a wide range of functions. Every version comprises the following: data import, images screening, intensity distribution histograming, data formatting changes, development of data layers on the basis of existed ones, image clustering, determination of classes features, best spectral zones depiction, classification of the chosen image or its unit, classification of results output, spectral brightness graphic output, colour layout of the chosen class correlation with spectral zones of initial image, supplementary utilities, including export of pixels brightness values for processing in other packages, etc.

Colour image synthesis means that each picture in three shooting channels has its own colour. RGB system is applied. Depending on the colour related ranges, a great deal of colour synthesized combinations of shooting channels can be specified. Each of these combinations enables its applicants to acquire particular information about surface objects and depict its ecological features [2].

Such values as 0,4-0,5; 0,5-0,6 and 0,6-0,75 mcm (that is, in the example with multispectral image Landsat 7 images 1-2-3 are presented) are spectral zones (of blue, green and red colour correspondingly) the most frequently used for synthesizing. This type of synthesis is a standard one (natural colour transfer).

Picture 1 demonstrates examples of image synthesis with distorted colour transfer. Type of synthesis 5-4-3 (picture 1a) presents a decoder much information and colour contrasts. Healthy vegetation has bright green colour, soils – pinky purple. This combination enables an applicant to analyze the cultivated lands, it is also very convenient for analyzing flora and forest communities.

In combination 4-3-2 (picture 1b) flora is red, city buildings are green and blue, soils colour varies from dark to light brown. Ice, snow and clouds are white or light blue (ice and clouds are presented on the edges). Pine woods will be of dark red or even brown colour in comparison with leaf bearing ones. This combination is rather popular and is used primarily for flora study, drainage and soil mosaic monitoring, as well as agricultural studies. In general, bright red

colours serve as an indicator of healthy and (or) broadleaf vegetation, whereas lighter colours characterize grassy vegetation or light forest/shrub vegetation.

Other types of synthesis applied for geological processes study, atmosphere features, water objects, smoke, forest cutting, soil humidity, etc. are applicable for ecological researches as well.
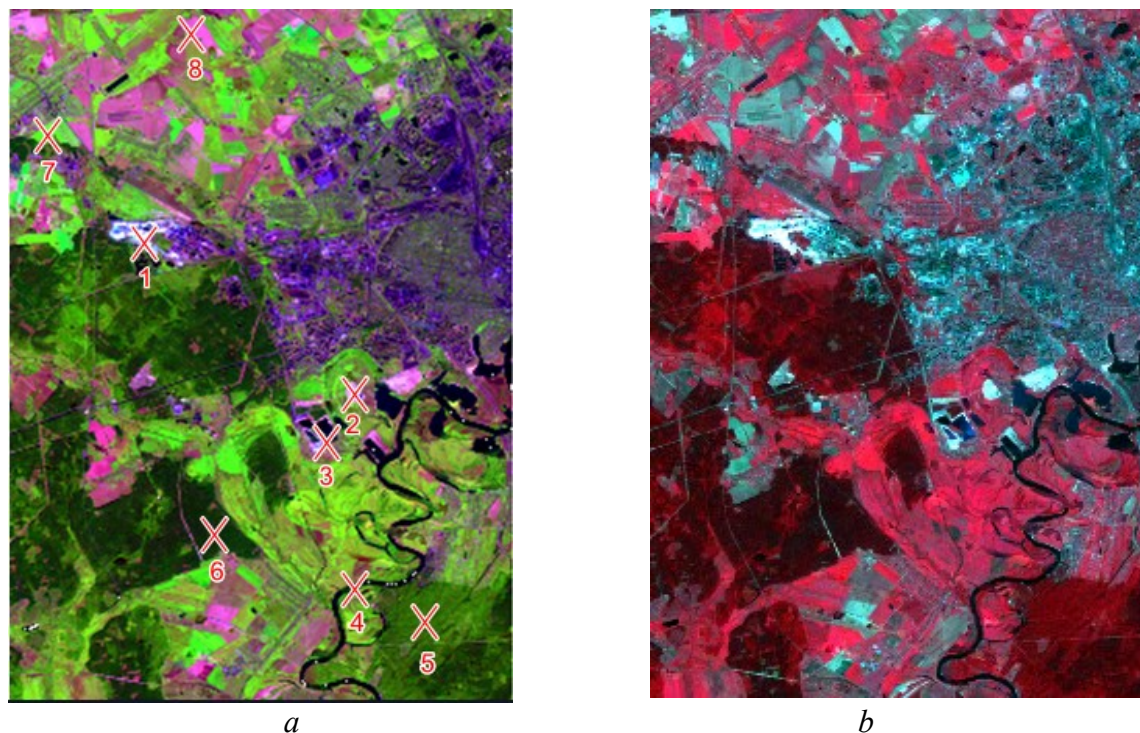

a   *b*

Fig 1 – Colour image synthesis of image Landsat 7 (western part of Gomel and Gomel suburbs): a – combination 5–4–3; b – combination 4–3–2

The third stage is dedicated to mastering methods of objects depiction on images by means of its decoding features. Residential areas, cultivated lands, forests, water objects, meadows and other natural and anthropogenic objects can be distinguished on the images [3]. Setting of images with maps of administrative territorial or natural (landscape, river estuaries, geomorphological objects) divisions makes it possible to find out proportion of land types within the limits of separate territorial divisions, evaluate its ecological state, make comparison of various territories and depict correlation between natural territorial features, level of anthropogenic influence on this territory and its ecological state. This stage also presupposes mastering of opportunities of acquisition vector shape-files of administrative division units of various levels of all countries (web-site: https://gadm.org) and creation of natural zone division files by means of GIS-packages.

Besides decoding features, graphs of spectral brightness – graphic depiction of correlation between wave length and reflex coefficients of the analyzed object (picture 2) – are used for decoding of objects and its features on space images. Multispectral image ranges numbers (levels of brightness are determined by means of MultiSpec programme) are indicated on X axis. Brightness characteristics in the considered spectral zones are indicated on the scale from 0 to 255 on Y axis. Each of surface objects is characterized by individual brightness in various ranges. Graphs of spectral brightness drawn for areas of eight natural and anthropogenic objects (picture 1a, the objects are marked and numerated) are presented as an example.

Graphs of spectral brightness enable to decode precisely various types of objects, even the visually identical ones. So, the following peculiarities of objects spectral brightness can be

specified for the presented example. Level of brightness in optical range channels for curve of phosphogypsum dumps spectral brightness of Gomel chemical factory is maximum among all the analyzed objects. Then, with the growth of wave length taken into consideration, it declines abruptly in infrared area of spectrum, brightness is the lowest in the sixth channel. As for the sand pit, brightness is the highest in the fifth channel. In general, shape of the graph is the same as previously, but the brightness in all channels is lower. Besides, for sand pit, unlike phosphogypsum dumps, brightness in the fifth channel is higher than in the fourth one.
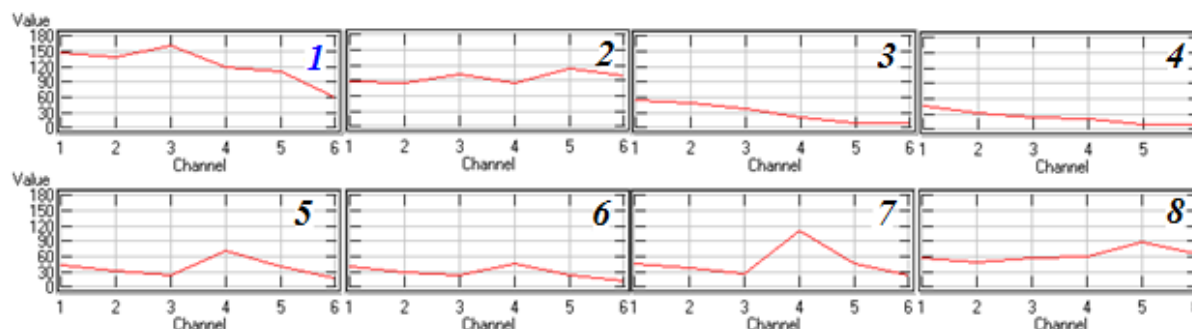

Fig 2 – Graphs of spectral brightness of natural and anthropogenic objects: 1– phosphogypsum dumps, 2 – sand pit, 3– water with suspensions, 4 – pure water, 5 – broadleaved forests, 6 – pine woods, 7 – unharvested cultivated lands, 8 – harvested cultivated lands

The minimum brightness is observed for water objects and declines with the growth of wave length. However, for polluted water objects brightness in all channels of visible range is higher than for pure ones. Graphs for pine forests differ from leaf bearing ones in brightness characteristics in the channels 4 and 5 (brightness is higher for leaf bearing forests). Curve of harvested cultivated land is of a similar character of curve of unharvested cultivated land, but there is no peak in the fifth one and brightness increases in the first, third, fifth and sixth channels.

The forth stage is dedicated to mastering methods of creation of derived images on the basis of space survey. Classification maps, clustering maps and characteristics distribution maps calculated by means of pixel brightness characteristics use in every channel are considered to be such derived images.

Classification map is the result of automated shots decoding by means of «studying» method, that is, initially a researcher depicts sample areas of the objects reflected on the image and then the programme classifies the other images pixels comparing it with the samples.

Clustering map is the result of automated decoding «without studying». In this case the programme divides all the pixels in the given class number itself with the application of ISODATA algorithm.

Knowledge of connection of vegetation structure and state relation with its reflective features enables to use space images for identification of vegetation types and its state [4]. For this purpose vegetation index is used – a criterion calculated as the result of operations with various spectral ranges (channels) of distant sounding data and related to vegetation characteristics in the given image pixel. Index effectiveness is determined by reflective peculiarities; these indices are of empirical nature.

Currently there are about 160 existing types of vegetation indices which are drawn through experiments (with peculiarities of vegetation and soils spectral reflectivity curves taken into consideration).

Calculation of the majority of vegetation indices is based on two more static (disregarding the other factors) lines of vegetation spectral reflectivity curve. Maximun chlorophyll solar radiation absorption accounts for red spectral zone (0,62 - 0,75 mcm) and maximum cell leaf

structure energy reflection – for near infrared zone (0,75 - 1,3 mcm), that is, high photosynthetic activity (as a rule, regarding big vegetation bio mass) determines lower reflection coefficients in spectral red zone and higher coefficients in near infrared zone. As known, ratio between these indicators enables to distinguish vegetation from other natural objects.

NDVI (Normalized Difference Vegetation Index) is the vegetation index most frequently used. It takes positive values for vegetation and the bigger green bio mass is, the higher this index becomes. NDVI is calculated by the following formula:

$$NDVI = \frac{NIR - ¿}{NIR + ¿}$$

where NIR stands for near infrared range radiation intensity (for Landsat 7 satellite images the range is 760-900 nm), RED – for red range radiation intensity (630-690 nm).

As an example a map of NDVI vegetation index distribution on the territory was drawn (picture 1). For this purpose New Channel from General Algebraic Transformation option of MultiSpec programme was used. Results visualization (picture 3) with pixels ranging by index value was made in Global Mapper GIS.

Fig 3 – Territorial distribution of vegetation index NDVI: 1 – less than 0; 2 – 0-0,2; 3 – 0,2-0,3; 4 – 0,3-0,4; 5 – 0,4-0,5; 6 – 0,5-0,6; 7 – more than 0,6.

Vegetation index mapping enables the user to depict ecosystem areas with high bio mass value, which is reflected in high index value.

IV. CONCLUSION

Ecology-related education of various topics, as well as use of land and water resources, is determined by demand for qualified teachers, specialists and ecologically educated leading staff members in the sphere of management, use and protection of natural resources. Teachers` continuing education in accordance with modern educational programmes, including GIS-technologies application in study process and research activities, is of primary importance as well.

REFERENCES

[1] Moiseeva L.V. Theoretical and methodological principles of ecological pedagogics // Uspehi sovremennogo estestvoznaniya (Achievements in the sphere of nowadays natural science). – 2004. – №7. – p. 67-69.
[2] Shovengerdt R. Distant sounding. Methods and models of image processing. – Moscow: Technosfera, 2013. – 592 p.
[3] Labutina I.A. Decoding of aerospace images. – Moscow: ASPECT-PRESS, 2004. –184 p.
[4] Cherepanov A.S. Vegetation indices // Geomatika. – № 2. – 2011. – p. 98-102.

Models and Methods of Evaluation of Information Sufficiency for Determining the Software Complexity and Quality Based on the Metric Analysis Results

Tetiana O. Hovorushchenko

Abstract— The aim of this study is the development of the models and methods for evaluating the information sufficiency for determining the software complexity and quality based on the metric analysis results. In this paper, the models and methods based on the comparative analysis of base ontology of subject area and ontology of concrete software are developed. The developed models and methods provide the sorting of all indicators, that absent in the software requirements specification (SRS), in descending the weights values, i.e. to prioritize additions in SRS.

Keywords—Information sufficiency, metric analysis, ontologies, software complexity, software quality

I. INTRODUCTION

Analysis of [1-5] revealed the fact that the causes of many software incidents are rooted in the SRS. The software quality is the degree of satisfaction of users or the degree of compliance to customers' needs [6-8]. Then, if the project objectives at the early lifecycle stages don't meet the needs of users, the software will not have high quality. Therefore, the quality and success of software project implementation significantly depend on the SRS and on the sufficiency of information in it. The *sufficiency of information* is the rational information saturation that eliminates information incompleteness (lack of necessary information).

Today the evaluation of indicators for the software quality and complexity metrics is conducted only at the stage of the quality evaluation for the ready source code [5]. But the SRS have all indicators, which are needed to the metrics calculation [5]. So the information sufficiency (as presence in the SRS all necessary indicators for metrics calculation) for future definition of the software complexity and quality can be evaluated on the basis of the SRS. And if some indicators are absent, then the SRS has insufficient information for metrics calculation and the developers have to make the necessary adjustments in the SRS.

The evaluation of sufficiency of the SRS information (presence in the SRS all necessary indicators for metrics calculation) provides the choice of software project in terms of its predicted quality and complexity at the early lifecycle stages, increases efficiency of project management due the validity of decisions, reduces the time of decision-making, reduces the costs for collection and processing of information at the later lifecycle stages (for example, during the software quality audit stage). The insufficiency of SRS information reduces the effectiveness and veracity of evaluating the software quality and complexity.

The *actual task* is the evaluating the sufficiency of the SRS information - for example, the possibility of calculating the values of the metrics of the software complexity and quality based on available indicators in SRS. So *the aim of this study* is the development of the models and methods for evaluating the information sufficiency for determining the software complexity and quality based on the metric analysis results.

T. O. Hovorushchenko, Khmelnytskyi National University, Khmenlytskyi, Ukraine (e-mail: tat_yana@ukr.net).

II. Formalized and Ontological Models of the Software Complexity and Quality Based on the Metric Analysis

During the analysis of software metrics as sources of information on its characteristics, the presence of cross-correlation of metrics was revealed because they have some joint indicators. The models of the software quality and complexity based on the metric analysis are necessary to develop for evaluating the correlation and the mutual influences of metrics and their indicators. In [9] it was proved that the software quality at the design stage (QDS) depends on 14 metrics, and software complexity at the design stage($CXDS$) depends on the 10 metrics with exact or predicted values: $QDS = \psi(sqm_1,...,sqm_{14})$, $CXDS = \phi(scxm_1,...,scxm_{10})$.

The set of software quality metrics at the design stage is: $SQM = \{Chp, Cpp, Rup, Mmt, Mbq, Sct, Sdt, Scc, Sqc, Cpt, Ccc, Fp, Lc, Dp\}$, where Chp – cohesion metric, Cpp – coupling metric, Rup – metric of the global variables calling, Mmt – time of models modification, Mbq – quantity of found bugs during the models inspection, Sct – software design total time, Sdt – design stage time, Scc – software design expected cost, Sqc – software quality audit expected cost, Cpt – software realization productivity, Ccc – code realization expected cost, Fp – functional points, Lc – effort applied by Boehm's model, Dp – expected development time by Boehm's model.

The set of software quality metrics at the design stage can be presented in the form of

$$SQM = \{SQM_{exv}, SQM_{prv}\} = \begin{Bmatrix} \{Chp, Cpp, Rup, Mmt, Mbq\}, \\ \{Sct, Sdt, Scc, Sqc, Cpt, Ccc, Fp, Lc, Dp\} \end{Bmatrix}, \text{ where } SQM_{exv} -$$

subset of software quality metrics with the exact values at the design stage, SQM_{prv} – subset of software quality metrics with the predicted values at the design stage.

The set of software complexity metrics at the design stage is: $SCXM = \{Is, N_{ZV}, MP, I, LOC_{оч}, HDiff, V(G), cl, N_{прогн}, Cmp\}$, where Is – Chepin's metric, N_{ZV} – Jilb's metric (absolute), MP – McClure's metric, I – Kafur's metric, LOC_{ep} – expected Lines Of Code, $HDiff$ – Halstead's metric, $V(G)$ – McCabe's metric, cl – Jilb's metric (logical), N_{ep} – expected quantity of program statements, Cmp – expected estimate of interfaces complexity.

The set of software complexity metrics at the design stage can be presented in the form of $SCXM = \{SCXM_{exv}, SCXM_{prv}\} = \{\{Is, N_{ZV}, MP, I\}, \{LOC_{ep}, HDiff, V(G), cl, N_{ep}, Cmp\}\}$, where $SCXM_{exv}$ – subset of software complexity metrics with the exact values, $SCXM_{prv}$ – subset of software complexity metrics with the predicted values at the design stage.

Then the models of software quality and complexity on the basis of metric analysis:

$$QDS = \psi(Chp, Cpp, Rup, Mmt, Mbq, Sct, Sdt, Scc, Sqc, Cpt, Ccc, Fp, Lc, Dp), \qquad (1)$$

$$CXDS = \phi(Is, N_{ZV}, MP, I, LOC_{ep}, HDiff, V(G), cl, N_{ep}, Cmp). \qquad (2)$$

Each of these metrics is a function of several indicators, moreover, quality and complexity metrics depend on 72 indicators, but only on 42 different indicators, then set of indicators of the software quality and complexity for further metric analysis has the form $SQCXI = \{sqcxi_1,...,sqcxi_{42}\}$.

The set of indicators of the software quality for further metric analysis has the form $SQI = \{sqi_1,...,sqi_{24}\}$ ($SQI \in SQCXI$), because the software quality metrics depend on 39 indicators, but only on 24 different indicators. The set of indicators of the software complexity for further metric analysis has the form $SCXI = \{scxi_1,...,scxi21\}$ ($SCXI \in SQCXI$),because the software complexity metrics depend on 33 indicators, but only on 21 different indicators (there are indicators that affect both quality metrics and complexity metrics, therefore they are both in the set SQI and in the set $SCXI$).

The models of software quality metrics have the form:

$$Chp = \psi_1(Cam, Iaom), \qquad (3)$$

where Cam − cohesion of actions in module, $Iaom$ − importance of actions order in module;

$$Cpp = \psi_2(Tmid, Tmopd, Pcd, fp, gp), \qquad (4)$$

where $Tmid$ − type of module input data, $Tmopd$ − type of module output data, Pcd − presence of common data, fp − quantity of preceding modules, gp − quantity of following modules;

$$Rup = \psi_3(Aup, Pup) = \frac{Aup}{Pup}, \qquad (5)$$

where Aup − quantity of real access to global variables, Pup − quantity of potential access to global variables;

$$Mmt = \psi_4(Qcl, Pd, Sdslc), \qquad (6)$$

where Qcl − quantity of code lines, Pd − project duration, $Sdslc$ − share of design stage;

$$Mbq = \psi_5(Qbm, Q_m), \qquad (7)$$

where Qbm − quantity of bugs of module, Q_m − quantity of modules;

$$Sct = \psi_6(Qcl, Pd); \qquad (8)$$

$$Sdt = \psi_7(Qcl, Pd, Sdslc); \qquad (9)$$

$$Scc = \psi_8(Qcl, Col) = Qcl \cdot Col, \qquad (10)$$

where Col − cost of one line;

$$Sqc = \psi_9(Svvtqlc, Sqavvtq, Qcl, Col), \qquad (11)$$

where $Svvtqlc$ − share of VVTQ stage, $Sqavvtq$ − share of quality audit in VVTQ;

$$Cpt = \psi_{10}(Qcl, Pd) ; \tag{12}$$

$$Ccc = \psi_{11}(Qcl, Col, Srslc) , \tag{13}$$

where $Srslc$ − share of realization stage in lifecycle;

$$Fp = \psi_{12}(EI, EO, EIN, ILF, ELF) , \tag{14}$$

where EI − quantity of external inputs, EO − quantity of external outputs, EIN − quantity of external requests, ILF − quantity of internal logic files; ELF − quantity of external logic files;

$$Lc = \psi_{13}(Qcl, Pt) = a \cdot Qcl^b , \tag{15}$$

where Pt − project type, which determines the COCOMO coefficients a, b;

$$Dp = \psi_{14}(Qcl, Pt) = c \cdot a \cdot Qcl^{b \cdot d} , \tag{16}$$

where Pt − project type, which determines the COCOMO coefficients a, b, c, d.

Thus, *the model of software quality based on the metric analysis (at the design stage)*:

$$QDS = \psi \begin{pmatrix} \psi_1(Cam, Iaom), \psi_2(Tmid, Tmopd, Pcd, fp, gp), \dfrac{Aup}{Pup}, \\ \psi_4(Qcl, Pd, Sdslc), \psi_5(Qbm, Q_m), \psi_6(Qcl, Pd), \\ \psi_7(Qcl, Pd, Sdslc), Qcl \cdot Col, \psi_9(Svvtqlc, Sqavvtq, Qcl, Col), \\ \psi_{10}(Qcl, Pd), \psi_{11}(Qcl, Col, Srslc), \\ \psi_{12}(EI, EO, EIN, ILF, ELF), a \cdot Qcl^b, c \cdot a \cdot Qcl^{b \cdot d} \end{pmatrix} . \tag{17}$$

As seen from formulas (3)-(16), some functions for calculating the software quality metrics are known (functions $\psi_3, \psi_8, \psi_{13}, \psi_{14}$), the remaining functions are uncertain.

The models of software complexity metrics have the form:

$$Is = \phi_1(P, M, C, T, Q_m) = Q_m \cdot (P + 2M + 3C + 0{,}5T) , \tag{18}$$

where P − quantity of variables for calculations and output, M − quantity of modified or created variables, C − quantity of control variables, T − quantity of not used variables;

$$N_{ZV} = \phi_2(Q_m, Q_{lem}) = Q_m \cdot Q_{lem} , \tag{19}$$

where Q_{lem} − quantity of links of each module;

$$MP = \phi_3(Q_m, fp, gp, X(Pm), Y(Pm)) = \sum_{i=1}^{Qm} (fp \cdot X(Pm) + gp \cdot Y(Pm)) , \tag{20}$$

where $X(Pm)$ – quantity of calls to module Pm, $Y(Pm)$ – quantity of calls from module Pm;

$$I = \phi_4(Q_m, W, R, WrRd) = Q_m \cdot (W \cdot R + W \cdot WrRd + WrRd \cdot R + WrRd \cdot (WrRd - 1)), \quad (21)$$

where W – quantity of procedures to update data structure, R – quantity of procedures to read from data structure, $WrRd$ – quantity of procedures to read and update data structure;

$$LOC_{ep} = \phi_5(Qcl) = Qcl; \quad (22)$$

$$HDiff = \phi_6(Qcl, NUOprtr, NUOprnd, NOprtr, NOprnd) = \frac{NUOprtr}{2} \cdot \frac{NOprnd}{NUOprnd}, \quad (23)$$

where $NUOprtr$ – quantity of unique operators, $NUOprnd$ – quantity of unique operands, $NOprtr$ – total quantity of operators, $NOprnd$ – total quantity of operands (depend on Qcl);

$$V(G) = \phi_7(E, N, NOprtr) = E - N + 2, \quad (24)$$

where E – quantity of control transfers, N – quantity of computing operators and expressions (depend on total quantity of operators $NOprtr$);

$$cl = \phi_8(NOprtr, L_{IF}, L_{LOOP}) = \frac{L_{IF} + L_{LOOP}}{NOprtr}, \quad (25)$$

where L_{IF} – quantity of logic operators, L_{LOOP} – quantity of cycle operators;

$$N_{ep} = \phi_9(NOprtr, Qcl); \quad (26)$$

$$Cmp = \phi_{10}(NOprnd, NUOprnd, Q_m). \quad (27)$$

As seen from formulas (18)-(27), some functions for calculating the software complexity metrics are known (functions $\phi_1 - \phi_8$), the remaining functions are uncertain.

Thus, *the model of software complexity based on the metric analysis (at the design stage)*:

$$CXDS = \phi \begin{pmatrix} Q_m \cdot (P + 2M + 3C + 0,5T), Q_m \cdot Q_{lem}, \sum_{i=1}^{Qm} (fp \cdot X(Pm) + gp \cdot Y(Pm)), \\ Q_m \cdot (W \cdot R + W \cdot WrRd + WrRd \cdot R + WrRd \cdot (WrRd - 1)), Qcl, \\ (NUOprtr / 2) \cdot (NOprnd / NUOprnd), (E - N + 2), \frac{L_{IF} + L_{LOOP}}{NOprtr}, \\ \phi_9(NOprtr, Qcl), \phi_{10}(NOprnd, NUOprnd, Q_m) \end{pmatrix}. \quad (28)$$

The models of the software quality and complexity based on the metric analysis show that there are indicators, which affect more than one metric. Thus, there is the metrics correlation

by some indicators. The existence of relationships between metrics affect their significance and weight [10], therefore should identify joint indicators for the metrics and should determine the significance (probability) of the indicators with the purpose of improving the veracity of the evaluations of the software quality and complexity. The knowledge of experienced professionals about the mutual influences and correlation of metrics are valuable in identifying the joint indicators, so they should be stored and used. The ontologies were selected for this knowledge reflection and accumulation.

The ontological model of software quality based on the metric analysis has the form: $O_{Q_{metr}} = <X_{Q_{metr}}, RX_{Q_{metr}}, F_{Q_{metr}}>$, where $X_{Q_{metr}}$ − finite set of metrics and indicators of the software quality, $RX_{Q_{metr}}$ − finite set of relationships between concepts, $F_{Q_{metr}}$ − finite set of interpretation functions for the software quality metrics and indicators.

Considering the model of software quality based on the metric analysis, the set of metrics and indicators of the software quality is:

$$X_{Q_{metr}} = \{SQM, SQI\} = \{x_{Q_{metr_1}}, ..., x_{Q_{metr_{38}}}\}, \qquad (29)$$

where $\{x_{Q_{metr_1}}, ..., x_{Q_{metr_{14}}}\} \in SQM$, i.e. $\{x_{Q_{metr_1}}, ..., x_{Q_{metr_{14}}}\} = \{sqm_1, ..., sqm_{14}\}$, $\{x_{Q_{metr_{15}}}, ..., x_{Q_{metr_{38}}}\} \in SQI$, then $\{x_{Q_{metr_{15}}}, ..., x_{Q_{metr_{38}}}\} = \{sqi_1, ..., sqi_{24}\}$.

The set of relationships between concepts $RX_{Q_{metr}}$ consists from relationship «depends on», i.e. $RX_{Q_{metr}} = \{"depends \ on"\}$. The set $F_{Q_{metr}}$ of interpretation functions for metrics and indicators of the software quality consists from function for quality depending on the metrics and functions for quality metrics depending on the indicators, i.e. $F_{Q_{metr}} = \{f_{Q_{metr_1}}, ...f_{Q_{metr_{15}}}\} = \{\psi(), \psi_1(), ..., \psi_{14}()\}$.

Thus *the base ontological model of the software quality based on the metric analysis*:

$$O_{Q_{metr}} = \{sqm_1, ..., sqm_{14}, sqi_1, ...sqi_{24}, "depends \ on", \psi(), \psi_1(), ..., \psi_{14}()\}. \qquad (30)$$

The ontological model of the concrete software quality based on the metric analysis:

$$O_{Q_{metr_{real}}} = \{sqm_1, ..., sqm_{nqm}, sqi_1, ...sqi_{nqi}, "depends \ on", \psi(), \psi_1(), ..., \psi_{14}()\}, \qquad (31)$$

where nqm ($nqm \leq 14$) − quantity of software quality metrics, which can be calculated on the basis of the available indicators in the SRS of concrete software, nqi ($nqi \leq 24$) − quantity of quality indicators, which are available in the SRS of concrete software.

The ontological model of software complexity based on the metric analysis has the form: $O_{CX_{metr}} = <X_{CX_{metr}}, RX_{CX_{metr}}, F_{CX_{metr}}>$, where $X_{CX_{metr}}$ − finite set of metrics and indicators of the software complexity, $RX_{CX_{metr}}$ − set of relationships between concepts, $F_{CX_{metr}}$ − set of interpretation functions for the software complexity metrics and indicators.

Considering the model of software quality based on the metric analysis, the set of metrics and indicators of the software complexity is:

$$X_{CX_{metr}} = \{SCXM, SCXI\} = \{x_{CX_{met\eta}}, ..., x_{CX_{met\eta31}}\},$$ (32)

where $\{x_{CX_{met\eta}}, ..., x_{CX_{met\eta10}}\} \in SCXM$, i.e. $\{x_{CX_{met\eta}}, ..., x_{CX_{met\eta10}}\} = \{scxm_1, ..., scxm_{10}\}$,

$\{x_{CX_{met\eta11}}, ..., x_{CX_{met\eta31}}\} \in SCXI$, then $\{x_{CX_{met\eta11}}, ..., x_{Q_{met\eta31}}\} = \{scxi_1, ..., scxi_{21}\}$.

The set of relationships between concepts $RX_{CX_{metr}}$ consists from relationship «depends on», i.e. $RX_{CX_{metr}} = \{"depends\ on"\}$. The set $F_{CX_{metr}}$ of interpretation functions for metrics and indicators of the software complexity consists from function for complexity depending on the metrics and functions for complexity metrics depending on the indicators, i.e. $F_{CX_{metr}} = \{f_{CX_{met\eta}}, ... f_{CX_{met\eta11}}\} = \{\phi(), \phi_1(), ..., \phi_{10}()\}$.

Thus *the base ontological model of the software complexity based on the metric analysis*:

$$O_{CX_{metr}} = \{scxm_1, ..., scxm_{10}, scxi_1, ... scxi_{21}, "depends\ on", \phi(), \phi_1(), ..., \phi_{10}()\}.$$ (33)

The ontological model of the concrete software quality based on the metric analysis:

$$O_{CX_{metr_{real}}} = \{scxm_1, ..., scxm_{ncxm}, scxi_1, ... scxi_{ncxi}, "depends\ on", \phi(), \phi_1(), ..., \phi_{10}()\},$$ (34)

where $ncxm$ ($ncxm \leq 10$) – quantity of software complexity metrics, which can be calculated on the basis of the available indicators in the SRS of concrete software, $ncxi$ ($ncxi \leq 21$) – quantity of complexity indicators, which are available in the SRS of concrete software.

III. FORMALIZED AND ONTOLOGICAL MODELS OF SOFTWARE REQUIREMENTS SPECIFICATION (IN TERMS OF THE AVAILABILITY OF INDICATORS FOR SOFTWARE METRICS CALCULATION)

Considering the SRS structure according to ISO 29148 [11], the SRS can be represented in the following formalized form (in terms of the availability in it of indicators for software quality and complexity metrics calculation):

$$SRS_{metr} = <R1_{metr}, R2_{metr}, R3_{metr}, R4_{metr}, R5_{metr}>,$$ (35)

where $R1_{metr}$ – set of complexity and quality indiactors of section 1 of SRS, $R2_{metr}$ – set of indiactors of section 2 of SRS, $R3_{metr}$ – set of indiactors of section 3 of SRS, $R4_{metr}$ – set of indiactors of section 4 of SRS, $R5_{metr}$ – set of indiactors of section 5 of SRS.

Some indicators may be contained in section 1 "Introduction" of the SRS, some indicators may be contained in section 3 "Specific requirements", some indicators may be contained in section 5 "Supporting information" of the SRS.

The ontological model of the SRS (in terms of the availability of indicators for software complexity and quality metrics calculation) has the form: $O_{SRS_{metr}} = <X_{SRS_{metr}}, RX_{SRS_{metr}}>$, where $X_{SRS_{metr}}$ – finite set of the software complexity and quality indicators in the SRS, $RX_{SRS_{metr}}$ – finite set of relationships between concepts.

Thus *the model of the SRS (in terms of the availability of indicators for software complexity and quality metrics calculation)* has the form:

$$SRS_{metr_i} = \left\langle \begin{array}{l} \{Pup, Sdslc, Svvtqlc, Sqavvtq, Srslc, Pt\}, \\ \varnothing, \\ \left\{ \begin{array}{l} Cam, Iaom, Tmid, Tmopd, Pcd, fp, gp, Aup, Q_m, EI, EO, EIN, \\ ILF, ELF, P, M, C, T, Q_{lem}, X(Pm), Y(Pm), W, R, WrRd \end{array} \right\}, \\ \varnothing, \\ \left\{ \begin{array}{l} Qcl, Pd, Qbm, Col, NUOprtr, NUOprnd, NOprtr, NOprnd, N, \\ L_{IF}, L_{LOOP} \end{array} \right\} \end{array} \right\rangle. \quad (36)$$

Considering the model of the SRS, the set of the indicators:

$$X_{SRS_{metr}} = \{SRS_{metr}, SQCXI\} = \{x_{SRS_{metr_1}}, ..., x_{SRS_{metr_{47}}}\}, \quad (37)$$

where $\{x_{SRS_{metr_1}}, ..., x_{SRS_{metr_5}}\} = \{R1_{metr}, ... R5_{metr}\}$, $\{x_{SRS_{metr_6}}, ..., x_{SRS_{metr_{47}}}\} = \{sqcxi_1, ..., sqcxi_{42}\}$. The set of relationships between concepts $RX_{SRS_{metr}}$ consists from relationship «contained in», i.e. $RX_{SRS_{metr}} = \{"contained\ in"\}$.

Thus *the base ontological model of the SRS (in terms of the availability of indicators for software complexity and quality metrics calculation)* has the form:

$$O_{SRS_{metr}} = \{R1_{metr}, ..., R5_{metr}, sqcxi_1, ... sqcxi_{42}, "contained\ in"\}. \quad (38)$$

The ontological model of the SRS of concrete software (in terms of the availability of indicators for software complexity and quality metrics calculation) has the form:

$$O_{SRS_{metr_{real}}} = \{R1_{metr}, ..., R5_{metr}, sqcxi_1, ... sqcxi_{ni}, "contained\ in"\}, \quad (39)$$

where ni ($ni \leq 42$) – quantity of complexity and quality indicators, which are available in the SRS of concrete software.

IV. ONTOLOGICAL METHODS OF EVALUATION OF INFORMATION SUFFICIENCY FOR DETERMINING THE SOFTWARE COMPLEXITY AND QUALITY BASED ON THE METRIC ANALYSIS RESULTS

Foremost, *the ontological method of evaluation of information sufficiency for determining the software complexity and quality based on the metric analysis results* was developed [12]. The base ontology for the subject domain "Software Engineering" (part "The software quality and complexity. Metric Analysis") was developed in [12]. For this ontology 4 software characteristics were selected: software project complexity, software complexity, software project quality, software quality. These characteristics are calculated on the basis of metrics, which in turn are based on indicators, according to the above models. The concept of the base ontology for the subject domain "Software Engineering" is shown on Figure 1.

Fig. 1 Concept of the base ontology for the subject domain "Software Engineering" (part "The software quality and complexity. Metric Analysis")

In evaluating the software complexity and quality should focus on those indicators that are part of multiple metrics simultaneously. By analogy with the method of evaluation of weights of software quality measures [13] let's evaluate the weights of SRS indicators, which are necessary for metrics calculation.

For evaluation of the weight of g-th indicator let's use the next formula:

$$\omega_{m_g} = \frac{k_{metr_{ind_g}}}{k_{ind}}, \tag{40}$$

where $k_{metr_{ind_g}}$ − quantity of metrics, which depend on g-th indicator; k_{ind} − total quantity of indicators (analysis of the above models showed that nowadays metrics of complexity and quality depend on 72 indicators, but on 42 different indicators, i. e. today $k_{ind} = 42$).

The developed models of the software complexity and quality, the base ontology for the subject domain "Software Engineering" (part "The software quality and complexity. Metric Analysis") provide the conclusions about the indicators, which are used for calculation of more than one metric, and about the quantities of metrics, which depend on each indicator (numerator of weights) [12]. In evaluating the quality and complexity metrics it's important to satisfy the availability in the SRS of those indicators, which have larger weights, with the purpose of providing the appropriate level of evaluations veracity.

The weighted ontology of the subject domain "Software Engineering" (part "The software quality and complexity. Metric Analysis") will be called the ontology, in which the complexity and quality indicators have the weights with the purpose of recommendations about the further satisfaction of these indicators in the SRS.

The method of evaluation of information sufficiency for determining the software complexity and quality based on the metric analysis results using the weighted ontology consists from next stages:

1) development of the weighted base ontology for the subject domain "Software Engineering" (part "The software quality and complexity. Metric Analysis");

2) analysis of the sections of SRS of concrete software for the availability of the indicators, which are necessary for metrics calculation, i.e. for the availability of the elements of set $X_{SRS_{metr}} = \{SRS_{metr}, SQCXI\} = \{x_{SRS_{metr_1}}, ..., x_{SRS_{metr_{47}}}\}$; generation and filling the template of ontology for concrete software, i.e. generation and filling the template of ontology $O_{SRS_{metr_{real}}} = \{R1_{metr}, ..., R5_{metr}, sqcxi_1, ...sqcxi_{ni}, "contained \ in"\}$;

3) comparing the developed weighed ontology for concrete software with the weighted ontology of the subject domain "Software Engineering" (part "The software quality and complexity. Metric Analysis"), i.e. comparing the set of indicators $\{sqcxi_1, ..., sqcxi_{ni}\}$ from ontological model of the SRS of concrete software $O_{SRS_{metr_{real}}} = \{R1_{metr}, ..., R5_{metr}, sqcxi_1, ...sqcxi_{ni}, "contained \ in"\}$ with the appropriate sets

$\{sqi_1,...,sqi_{24}\}$, $\{scxi_1,...,scxi_{21}\}$ of the base ontological models of the software quality and complexity based on the metric analysis

$$O_{Q_{metr}} = \{sqm_1,...,sqm_{14}, sqi_1,...sqi_{24}, "depends\ on", \psi(), \psi_1(),...,\psi_{14}()\}$$ and

$$O_{CX_{metr}} = \{scxm_1,...,scxm_{10}, scxi_1,...scxi_{21}, "depends\ on", \phi(), \phi_1(),...,\phi_{10}()\};$$

4) identifying the indicators, which are absent in the weighed ontology for concrete software, i.e. forming set $\{sqcxi_1,...,sqcxi_{(42-ni)}\} = \{sqcxi_1,...,sqcxi_{42}\} \setminus \{swcxi_1,...,sqcxi_{ni}\}$,

where $\{sqcxi_1,...,sqcxi_{42}\} \in (O_{Q_{metr}} \cup O_{CX_{metr}})$, $\{sqcxi_1,...,sqcxi_{ni}\} \in O_{SRS_{metr_{real}}}$ (if these sets are not empty, then SRS information is not sufficient for calculating the metrics of software complexity and quality; the more elements are in these sets, the smaller sufficiency of SRS information is); sorting of the missing indicators in descending the values of weights; herewith the numerator of the weight of each missing indicator indicates the number of software metrics that cannot be calculated without this indicator;

5) identifying the metrics, which cannot be calculated on the basis of available indicators;

6) identifying the software characteristics, which cannot be calculated on the basis of the metrics, which can be calculated on the basis of available indicators;

7) making the decision on the need to supplement of the SRS by the indicators, if there are metrics and characteristics whose values can not be determined based on available indicators; herewith the indicators with larger weights (the first in the sorted list of missing indicators) should be added in the SRS first of all;

8) repeating the stages 2-7 until it will be possible to identify all the metrics and software characteristics, or until forming the conclusion about insufficient data for determining the software complexity and quality with high veracity degree.

On the basis of the base ontology of the subject domain "Software Engineering" (part "The software quality and complexity. Metric Analysis"), which is represented in [12], let's develop the weighted base ontology for the subject domain "Software Engineering" (part "The software quality and complexity. Metric Analysis"). In this weighted ontology, there is information about the weights of the SRS indicators, which are necessary for the metrics calculation. The parts of this weighted base ontology are: the weighted base ontology for the software project complexity (Figure 2), the weighted base ontology for the software complexity, the weighted base ontology for the software project quality, the weighted base ontology for the software quality (all these ontologies are similar to ontology on Figure 2 and are developed according to above models).

Fig. 2 The weighted base ontology for the software project complexity

Marking the weights of complexity and quality indicators in the weighted base ontology provides the sorting all the missing in the SRS indicators in descending values of weights, i.e. prioritizes their additions in the SRS.

V. Experiments

The SRS of the automated system for large-format photo print was analyzed, on the basis of this SRS the ontology for the concrete software was developed.

On Figure 3 the ontology for the complexity of concrete software project (the part of the ontology for the concrete software project) is presented. The metrics that cannot be calculated on the basis of the available in the SRS indicators are circled in Figure 3.

Fig. 3 The ontology for the complexity of concrete software project

Comparative analysis of the developed ontology for the automated system for large-format photo print with the base ontology of the subject domain "Software Engineering" (part "The software quality and complexity. Metric Analysis") provides the conclusion that 9 (from 42) indicators are absent in the developed ontology for the concrete software project, i.e. *the SRS information is insufficient for software metrics calculation* (20 metrics from 24 cannot be calculated). For example, the compare Fig. 2 and Fig. 3 provides the conclusion that: the SRS information is insufficient for calculating the Chepin's metric, McClure's metric, Kafur's metric, and in the SRS information for calculating the Jilb's metric (absolute) is at all absent.

The sorted list of the missing indicators in descending the values of weights is: 1) quantity of code lines – 12/42; 2) quantity of modules – 6/42; 3) project duration – 4/42; 4) total quantity of operators – 4/42; 5) cost of one line – 3/42; 6) project type – 2/42; 7) share of design stage in lifecycle – 2/42; 8) quantity of control variables – 1/42; 9) quantity of links of each module – 1/42. This sorted list indicates the priority of indicators and the consistency of their review and addition in the SRS.

VI. Conclusion

The metric analysis is an effective mean of evaluating the software complexity and quality on condition of the availability of sufficient information for this. One of the factors, which affect to the veracity of such information, is the sufficiency of the information in the SRS regarding the indicators for metrics calculation. So the development of models and methods of evaluation of information sufficiency for determining the software complexity and quality, in general, enhances the veracity of evaluates of the software complexity and quality.

The developed model of the software quality and complexity based on the metric analysis, formalized and ontological model of the SRS (in terms of the availability of indicators for

software complexity and quality metrics calculation) became the basis for the development of the ontological methodology of complex evaluation of the software quality and complexity.

The analysis of the software metrics as sources of information on its characteristics, revealed the cross-correlation of these metrics because they have some joint indicators. The ontologies were selected for the reflection and accumulation of the knowledge of experienced professionals about the mutual influences and correlation of metrics. The ontologies became the basis of the ontological method of evaluation of information sufficiency for determining the software complexity and quality based on the metric analysis results. The correlation of metrics on indicators, that is displayed in the base ontology, taken into account in evaluating the weights of indicators. The lack of indicators, for which there is the correlation, can impair the accuracy and veracity of evaluations of the software complexity and quality. The correlation of metrics on some indicators increases the importance of these indicators in evaluating the software complexity and quality, thus increases the weights of indicators.

The developed method of evaluation of information sufficiency for determining the software complexity and quality based on the metric analysis results using the weighted ontology provides the conclusion about the insufficiency of the SRS information for metrics calculation, the sorting missing in the SRS indicators, prioritization their addition to the SRS.

REFERENCES

[1] C. Jones, O.Bonsignour, *The Economics of Software Quality*. Boston: Pearson Education, 2012, 588 p.

[2] T. Ishimatsu, N.G. Levenson, J. P. Thomas, C. H. Fleming, M. Katahira, Yu. Miyamoto, R. Ujiie, H. Nakao, N. Hoshino, "Hazard Analysis of Complex Spacecraft Using Systems-Theoretic Process Analysis," *Journal of Spacecraft and Rockets*, vol. 51, no. 2, pp. 509-522, 2014.

[3] E. Yourdon, *Death March: The Complete Software Developer's Guide to Surviving "Mission Impossible" Projects*. Prentice Hall, 2003, 256 p.

[4] C. Shamieh, *Systems Engineering for Dummies*. Wiley Publishing, 2011, 74 p.

[5] D. Maevskiy, Y. Kozina, "Where and When Is Formed of Software Quality?," *Electrical and Computer Systems*, no. 18, pp. 55-59, 2016. *(in Russian)*

[6] ISO/IEC 25010:2011. Systems and software engineering. Systems and software Quality Requirements and Evaluation (SQuaRE). System and software quality models. ISO/IEC, 2011, 34 p.

[7] ISO/IEC 25030:2007. Software engineering. Software product Quality Requirements and Evaluation (SQuaRE). Quality requirements. ISO/IEC, 2007, 36 p.

[8] ISO/IEC TR 19759:2015. Software Engineering. Guide to the software engineering body of knowledge (SWEBOK). ISO/IEC, 2015, 336 p.

[9] O. Pomorova, T. Hovorushchenko, "The Intelligent Decision Support System for Choice of Software Project," *Journal of Information, Control and Management Systems,* vol. 10, no.1, pp.87-96, 2012.

[10] S. Sugiyanto, S. Rochiman, "Integration of DEMATEL and ANP methods for calculate the weight of characteristics software quality based model ISO 9126," in *International Conference on Information Technology and Electrical Engineering,* Yogyakarta (Indonesia), 2013, pp.143-148.

[11] ISO/IEC/IEEE 29148-2011. Systems and software engineering. Life cycle processes. Requirements engineering. ISO/IEC/IEEE, 2011, 28 p.

[12] T. Hovorushchenko, O. Pomorova, "Method of Assessment of Information Sufficiency for Determination of Software Complexity and Quality Based on The Comparative Analysis of Ontologies," *Radioelectronic and Computer Systems*, no.6, pp. 59-68, 2016. *(in Ukrainian)*

[13] T. Hovorushchenko, O. Pomorova, "Evaluation of Mutual Influences of Software Quality Characteristics Based ISO 25010:2011," in *XI International Scientific Conference on Computer Sciences and Information Technologies*, Lviv (Ukraine), 2016, pp.80-83.

Classification by Ordered Fuzzy Decision Tree

Jan Rabcan, Makpal Zhartybayeva

Abstract— Nowadays, the classification represents a significant part of the data mining. The object of the classification is assigned to the new data sample the output property (class label) based on previous, learned experience. In this paper the approach of ordered fuzzy decision tree is considered. The fuzzy logic can reduce the uncertainness of initial data and it is closer to natural way of human thinking. Chosen classification model is evaluated by estimation of error and accuracy of the resulting classification.

Keywords—Classification, Fuzzy, Decision tree, Data mining.

I. INTRODUCTION

Data mining is the process of analysis of data from the various perspective and summarization the results on useful information [1]. The goal of data mining is to discover salutary knowledge stored in huge databases and repositories [1]. Data mining includes many techniques like prediction, classification, clustering, association rules, estimation and affinity grouping [2]. One of the mentioned data mining tasks is classification. The aim of the classification is to assign the class label for new instance. The popular way of classification is decision tree technique. Nowadays, there are many methods for induction of decision tree. One of the first decision tree technique has been published by J. R. Quinlan in [3]. The main idea of the ID3 algorithm is to choose the associate attribute to each node with minimal entropy or maximal information gain [3]. J. R. Quinlan modified ID3 algorithm in [4]. The modified version is called C4.5. Also C4.5 algorithm deals with information entropy. The splitting criterion is the normalized information gain [4]. Many real words problems are uncertainties and noisy. In this case, the crisp classification can be difficult to perform. The usage of fuzzy sets can be useful to describe real-world problems with higher accuracy [5] and more naturally to the way of human thinking. For this reason, the fuzzy decision trees are considered in this paper. At present time, many algorithms for induction of fuzzy decision tree have been proposed. One popular method has been described by Yuan and Shaw in [6]. The induction is based on the reduction of classification ambiguity with fuzzy evidence. Another way of FDT induction is based on fuzzy rules and published by Xianchang Wang in [7]. In contrast with "traditional" decision trees in which only a single attribute is taken into account at an each node, the node of the proposed decision trees in [7] involves a fuzzy rule which take into account multiple attributes. The next approach has been presented in [8]. This algorithm for Ordered Decision Tree (OFDT) is proposed in [8] and used for needs of this paper. OFDT algorithm takes only one attribute to each level of the decision tree. This feature can be considered as an advantage, because it allows constructing OFDT as a parallel process [5]. The criterion to choose attribute associated with given level is cumulative information estimations of fuzzy sets [9].

The usage of considered classification method has been evaluated on well know public dataset *Pima Indians Diabetes Database*. The dataset contains medical records of female patients and the goal is to estimate whatever a patient has signs of diabetes or not. Evaluation of OFDT is done by estimation of error and accuracy of the resulting classification.

II. FUZZY LOGIC

Fuzzy logic is one way to represent multi-valued logic. In classical Boolean logic, variables can reach only "crisp" values 1 or 0. Fuzzy logic describes variables using membership function

J. Rabcan, University of Zilina, Faculty of Management Science and Informatics, Zilina, Slovakia (e-mail: jan.rabcan@fri.uniza.sk)
M. Zhartybayeva, Faculty of Information Technology, L.N. Gumilyov Eurasian National University, Astana, Kazakhstan.

where variables are between 0 and 1 which represent degrees of membership [10]. Let set U is Universe of discourse. Universe of discourse contains all investigated samples. The fuzzy set A in set U has following definition. The set A is characterized by membership function as follow [11]:

1. $\mu_A(u) = 0$ if and only if x is not the member of set A
2. $\mu_A(u) = (0,1)$ if and only if x is not the full member of set A
3. $\mu_A(u) = 1$ if and only if x is the full member of set A

The function $\mu_A(u)$ is the membership function. This function for all elements in U assigns the value between 0 and 1. Fuzzy set is defined as an ordered set of pairs [11]:

$$A = \{(u, \mu_A(u)), u \in A\} \tag{1}$$

The following example illustrates usage of fuzzy logic on speed variable. If speed of car traveling on the highway is 100 km/h, it is hard to tell, if car is moving fast or not. In classical "crisp" logic the result is *low speed = 0, medium speed =1, high speed = 0*. In fuzzy logic the result is *low speed = 0, medium speed =0.6, high speed = 0.3*.

Fig. 1 Fuzzy set of the car movement speed

III. DATA

The repository for data storage is represented in tabular form (Table 1). The table consists of $n+1$ columns that correlate with n input attributes and 1 output attribute. The i-th column, for $i = 1,\ldots, n +1$, is divided into q_i sub-columns. The j_i-th sub-column, for $j_i = 1,\ldots, q_i$, corresponds to the j-th value of the attribute represented by the i-th column. Every row of the table represents one sample of initial (collected) data.

The proposed classification system deals with fuzzy attributes. Fuzzy attribute A_i is a linguistic attribute, and its possible values are also stored in the repository. Each value $A_{i_q} \in \langle 0,1 \rangle$ and $\sum_{q=1}^{q_i} A_q = 1$. These possibilities correspond to a membership function of fuzzy data [12]. These requirements on representation of initial data are caused by the method of FDT induction.

TABLE 1
TRAINING AND TESTING DATASETS

Input attributes, Ai								Output attribute, B		
A1			A2				An			
A11	A12	A13	A21	A22	· · · ·	An1	An2	B1	B2	B3
0.1	0.5	0.4	0.6	0.4	. . .	0.5	0.5	0.0	0.0	1.0
0.2	0.1	0.7	0.1	0.9	. . .	0.8	0.2	0.3	0.7	0.0
	. . .		. . .		. . .	. . .		. . .		
0.3	0.3	0.4	0.0	1.0	. . .	0.4	0.6	0.1	0.4	0.5

Repository (data collection) for the FDT induction.
Results of measurement, expert evaluation, or monitoring

The used dataset also contains numerical attributes. OFDT works only with fuzzy attributes. Therefore, all numerical attributes must be fuzzied. Yuan and Shaw have suggested a simple method of how to generate a set of membership functions for transforming numeric data to fuzzy data, which is described in [1]. Let attribute A be numeric attribute. The algorithm in first step divides values x_i of A into Q intervals $\{Q_1, \ldots, Q_q, \ldots, Q_Q\}$ by using arbitrary clustering algorithm. Every interval Q_q contains the center C_q. Then the membership functions are created based on the following rules. For the first linguistics term $x_{i,1}$, the next membership function is used [5]:

$$\mu_{x_{i,1}}(x) = \begin{cases} 1 & x \leq C_1 \\ \frac{C_2 - x}{C_2 - C_1} & C_1 < x < C_2 \\ 0 & x \geq C_2 \end{cases} \qquad (2)$$

Every linguistics term $x_{i,q}$, for $q = 2,3, \ldots, N-1$, has a membership function of the following form:

$$\mu_{x_{i,q}}(x) - \begin{cases} 0 & x \leq C_{j-1} \\ \frac{x - C_{j-1}}{C_j - C_{j-1}} & C_{j-1} < x \leq C_j \\ \frac{C_{j+1} - x}{C_{j+1} - C_j} & C_j < x \leq C_{j+1} \\ 0 & x \geq C_{j+1} \end{cases} \qquad (3)$$

The last term $x_{i,Q}$ has the membership function of the form of:

$$\mu_{x_{i,Q}}(x) = \begin{cases} 0 & x \leq C_{k-1} \\ \frac{x - C_{k-1}}{C_k - C_{k-1}} & C_{k-1} < x \leq C_k \\ 1 & x \geq C_k \end{cases} \qquad (4)$$

The following example illustrates the fuzzification method, where a numerical attribute represent the age of female. The investigated attribute is transformed into four groups: young, early adulthood, middle-aged and old age (Fig. 2). The first set young and the last set old have a form which can be described by four corners and the form is trapezoidal [1], while the sets between the first and the last set has triangular form. Therefore, they are described by three corners. For example, the set young is described by (0, 0, 16, 32), while the set early adulthood can be expressed using (16, 32, 48). Fuzzy set of investigated attribute (age) is shown on Fig. 2.
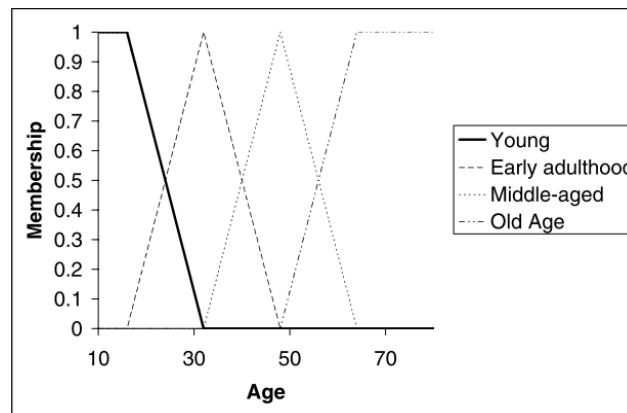

Fig. 2 Membership functions obtained after transformation

IV. ORDERED FUZZY DECISION TREE

An Ordered FDT (OFDT) is a specific type of decision trees, which has exactly one attribute in every level of the tree. The algorithm for OFDT induction has two main aspects. The first is choosing expandable attributes for each node, and the second is establishing the leaf nodes. Expandable attribute is chosen by cumulative mutual information [13]. Usage of this information characteristic allows parallel tree induction. The cumulative mutual information is estimated as follows:

$$\Delta \mathbf{I}\left(B; A_{i_1}, \ldots, A_{i_{q-1}}, A_{i_q}\right) = \mathbf{I}\left(B; A_{i_1}, \ldots, A_{i_{q-1}}, A_{i_q}\right) - \mathbf{I}\left(B; A_{i_1}, \ldots, A_{i_{q-1}}\right) \quad (5)$$

The attribute with the greatest information is chosen to associate with all nodes of the given level. The criterion for choosing expandable attributes of the OFDT can take into account the cost to measure value of attribute A. Then the criterion to choose attribute is following:

$$q = \operatorname{argmax}\left(\frac{\Delta \mathbf{I}\left(B; A_{i_1}, \ldots, A_{i_{q-1}}, A_{i_q}\right)}{Cost\left(A_{i_q}\right)}\right) \quad (6)$$

The OFDT algorithm has to establish leaf node during the induction phase. The presented algorithm uses a defined threshold values α and β. The threshold β represents the confidence level. The threshold α reflects the frequency of occurrences in the given node. Every internal node of the tree is declared as a leaf if at least one of the following conditions is satisfied:

$$f\left(U_{i_q j_q}\right) = \frac{M\left(A_{i_1 j_1} \times \ldots \times A_{i_q j_q}\right)}{N} \leq \alpha \quad (7)$$

$$2^{-\mathbf{I}\left(B | A_{i_1 j_1}, \ldots, A_{i_q j_q}\right)} \geq \beta \quad (8)$$

where $U_{i_q j_q} = \{A_{i_1, j_1}, \ldots, A_{i_1, j_1}\}$ means values of input attributes and $M(A_{i,j}) = \sum_{k=1}^{N} \mu_{Aj}(x_k)$. $\mathbf{I}\left(A_{i_1 j_1} | A_{i_2 j_2}\right) = \log_2 M\left(A_{i_1 j_1}\right) - \log_2 M\left(A_{i_2 j_2} \times A_{i_1 j_1}\right)$. Steps of the OFDT algorithm are described in Table 2.

TABLE 3
ALGORITHM FOR OFDT INDUCTION

0	• q is the current level. Put q $=$ 0.
	• Insert all input attributes into the set of unused attributes ua.
1	• Select an attribute from ua by $$\operatorname{argmax}\left(\frac{\Delta \mathbf{I}\left(B; A_{i_1}, \ldots, A_{i_{q-1}}, A_{i_q}\right)}{Cost\left(A_{i_q}\right)}\right)$$
2	• Associate the chosen attribute with each node in level q and remove it from ua.
3	• Check every node in level q, if node is a leaf.
4	• If all nodes in level q are leaf or ua is empty go to 5. • Else put q = q + 1 and recursively go to 1.
5	• Exit.

The threshold values affect on tree depth and size. Size of the tree presents number of nodes. The parameter β is the threshold value of the confidence level. If some node has bigger or equal confidence level as β, this node has to become a leaf. Increasing the value of beta causes an increase of the size of the tree. The parameter α represents the threshold of the minimal

frequency of a given branch. If some node has lower or equal frequency as α, this node has to become the leaf. The parameter α also affects the size of the tree. In this case, bigger α causes the smaller tree size. The threshold values must be set to satisfied value to perform accurate classification [3], [9]. One possible way, how to estimate threshold values is run the algorithm more times with different threshold values and then the best combination is chosen. The process is described by diagram on Fig. 2.

Fig. 2. Diagram of threshold values estimation

V. ANALYSIS OF FUZZY DECISION TREE

The trees are analyzed by error and accuracy of the classification. During the error estimation, the tree is built from the whole training set and all instances in the training set are classified. The result of error estimation is the ratio between classification mistakes and the number of instances in the dataset, which represents the percentage of incorrectly classified instances. In case of accuracy estimation, the training dataset is divided into two dataset. The bigger dataset contains 80 % of the divided training dataset and this dataset is used to build decision tree. The second dataset contains 20 % of the divided dataset and samples from this dataset are used for classification. The result of accuracy of classification is percentage of correctly classified instances.

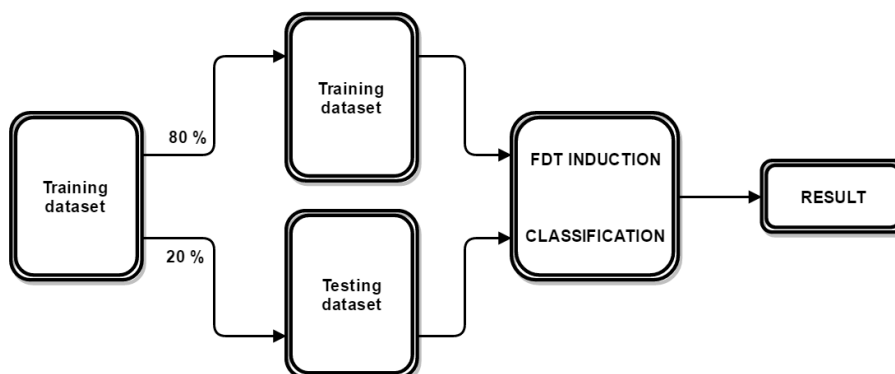Fig. 3 Diagram of classification error estimation

Fig. 4 Diagram of classification accuracy estimation

VI. USAGE OF ORDERED DECISION TREE

The usage of OFDT is demonstrated on the public dataset *Pima Indians Diabetes Database.* The dataset is available on [14]. Dataset contains medical records of diabetic patients. In particular, all patients here are females at least 21 years old of Pima Indian heritage. The population lives near Phoenix, Arizona, USA. The goal is investigate whether the patient shows signs of diabetes. The dataset has been gathered by National Institute of Diabetes and Digestive and Kidney Disease. The dataset contains 786 number of instance. All attributes are numerical. The count of input attributes is 8. The dataset contains also one numerical output attribute. The Table 2 contains attribute description and brief statistical analysis of attributes.

TABLE 4
ATTRIBUTE DESCRIPTION AND BRIEF STATISTICAL ANALYSIS OF ATTRIBUTES

Attribute	Mean	Standard Deviation
Number of times pregnant	3.8	3.4
Plasma glucose concentration a 2 hours in an oral glucose tolerance test	120.9	32.0
Diastolic blood pressure	69.1	19.4
Triceps skin fold thickness	20.5	16.0
2-Hour serum insulin	79.8	115.2
Body mass index	32.0	7.9
Diabetes pedigree function	0.5	0.3
Age (years)	33.2	11.8

The illustration of full OFDT is not suitable for this paper because the resulting tree consists of 39 nodes and its depth is 8. Hence, the threshold values have been chosen to induct a tree with suitable size and depth for visualization. Illustrated OFDT has following threshold values $\beta = 0.75$ and $\alpha = 0.25$. The classification error of this tree is 33.268 %. The accuracy of this example is 35.651 %. The tree is painted on Fig. 6. The optimal threshold values for classification error estimation are $\alpha = 0$ and $\beta = 1$. These values have been established by repeated runs of algorithm and the best combination has been chosen. In this case the tree is much robust, but classification error is only 0.036 %. The accuracy of classification is 0.947 %.
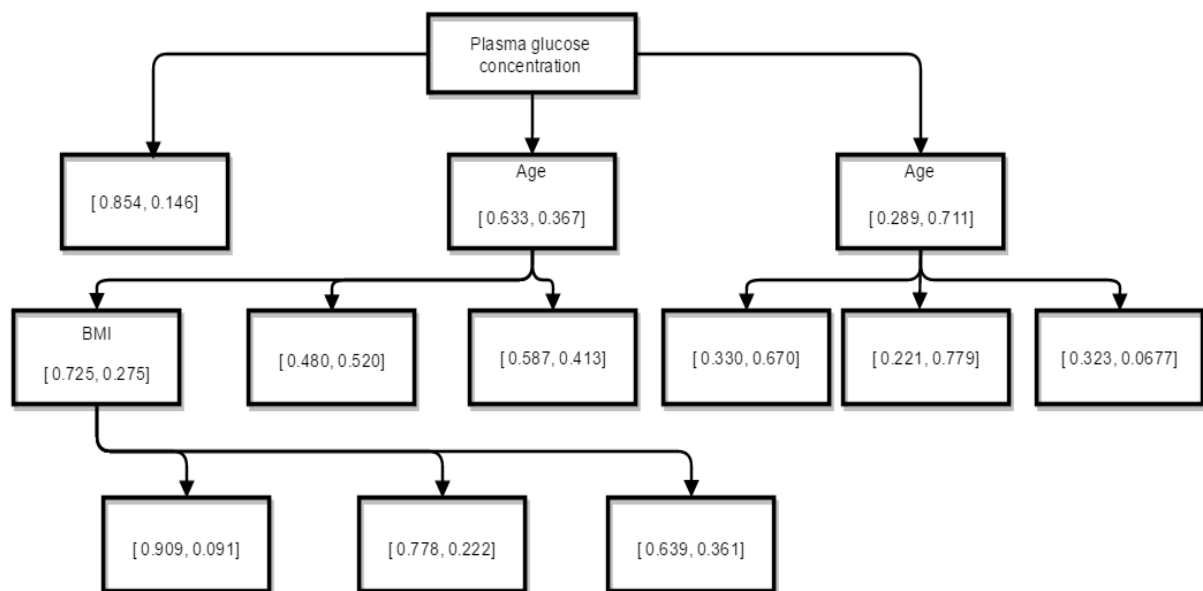

Fig. 5 Picture of small OFDT for Pima Indians Diabetes Database

VII. CONCLUSION

In this paper the classification method using approach of fuzzy decision trees is considered. In the paper the fuzzyification algorithm is also described. The presented algorithm OFDT is based on cumulative information estimations of initial data. The considered method is

demonstrated on the public dataset: *Pima Indians Diabetes Database.* Algorithm is able to classify new instance with satisfied accuracy is 94.7 %. The classification error is 0.036 %.

REFERENCES

[1]   J. I. Maletic and A. Marcus, *Data Mining and Knowledge Discovery Handbook*. 2005.

[2]   V. Levashenko, E. Zaitseva, M. Kvassay, and Deserno, "Reliability Estimation of Healthcare Systems using Fuzzy Decision Trees," *Ann. Comput. Sci. Inf. Syst.*, vol. 8, pp. 331–340, 2016.

[3]   J. R. Quinlan, "Induction of Decision Trees," *Mach. Learn.*, vol. 1, no. 1, pp. 81–106, 1986.

[4]   J.R. Quinlan, "Programs for machine learning. Morgan Kaufmann," p. C4. 5, 1993.

[5]   J. Rabcan, "Ordered Fuzzy Decision Trees Induction based on Cumulative Information Estimates and Its Application," *ICETA*, p. 6, 2016.

[6]   Y. Yuan and M. J. Shaw, "Induction of fuzzy decision trees," *Fuzzy Sets Syst.*, vol. 69, no. 2, pp. 125–139, 1995.

[7]   X. Wang, X. Liu, W. Pedrycz, and L. Zhang, "Fuzzy rule based decision trees," *Pattern Recognit.*, vol. 48, no. 1, pp. 50–59, 2015.

[8]   V. Levashenko, E. Zaitseva, and S. Puuronen, "Fuzzy Classifier Based on Fuzzy Decision Tree," *EUROCON 2007 - Int. Conf. Comput. as a Tool*, pp. 823–827, 2007.

[9]   V. Levashenko and E. Zaitseva, "Fuzzy Decision Trees in Medical Decision Making Support System," *Computer Science and Information Systems*. " in 2012 Federated Conference on Computer Science and Information Systems, pp. 213–219, 2012.

[10]   L. Biacino and G. Gerla, "Fuzzy logic, continuity and effectiveness," *Arch. Math. Log.*, vol. 41, no. 7, pp. 643–667, 2002.

[11]   Y. Yao, "A comparative study of fuzzy sets and rough sets," *Inf. Sci. (Ny).*, vol. 109, no. 1–4, pp. 227–242, 1998.

[12]   H. Tanaka, L. T. Fan, F. S. Lai, and K. Toguchi, "Fault-tree analysis by fuzzy probabilities," *Ieee*, vol. R-32, pp. 453–457, 1983.

[13]   V. G. Levashenko and E. N. Zaitseva, "Usage of New Information Estimations for Induction of Fuzzy Decision Trees," *Lect. Notes Comput. Sci.*, vol. 2412, pp. 493–499, 2002.

[14]   "Pima Indians Diabetes Database." [Online]. Available: https://archive.ics.uci.edu/ml/datasets/Pima+Indians+Diabetes.

Application of Markov Modeling for Safety Assessment of Self-Diagnostic Programmable Instrumentations and Control Systems

Valentyna Butenko, Oleg Odarushchenko, Vyacheslav Kharchenko, Viktoriya Moskalets, Elena Odarushchenko, Oleksii Strjuk

Abstract—Markov modeling is a well-known analytical state space modeling technique which is widely applied for quantitative analysis of safety-critical systems. There are few roadblocks for greater application of Markov modeling: accounting of additional system components increases the model state space and complicates analysis; the non-numerically sophisticated user may find it difficult to select method and tool to provide an accurate analysis of constructed Markov model. Thus, achieving highly trusted result for safety-critical systems is a nontrivial task. In this paper we present the case-study on application of Markov modeling with deep testing the model features, for safety analysis of industrial self-diagnostic, programmable FPGA-based Instrumentation and Control system which operates on Nuclear Power Plant.

Keywords—Instrumentation and Control system, Reactor Trip System, Markov model, metric-based approach

I. INTRODUCTION

An accurate safety assessment is a key task during development and certification of safetycritical systems as it allows to demonstrate that relevant requirements have been met. Detailed safety analysis is extremely important in case of design of Instrumentation and Control systems (I&Cs) that function on Nuclear Power Plants (NPP) because of potential risks for environment and people. To prevent any accident that may occur during continuous work of I&Cs various software and hardware diversity architectures were developed as well as deep self-diagnostic methods [1].

Normal operating and emergency protection systems are typical examples of NPP I&Cs.

The range of approaches were standardized to determine how to achieve the high accuracy goal. Among such widely applied techniques as FMEA, FMECA, FMEDA etc. stands Markov modeling. This is a well-known analytical state space modeling technique that was applied through decades to assess mainly dependability of safety-critical systems, meanwhile it is stated as one that can be applied for safety assessment as well [2]. With state space approaches the modeler can analyze failure/repair dependencies, shared repair facilities [3] results of errors in self-diagnostic tools and provide the detailed presentation of system behavior for communication with engineering team [4].

Modeling components interaction and interdependencies expands the model significantly, thus making the precise computation of system transient measures almost infeasible. Whilst numerical methods and imitation modeling can be applied to handling this problem, they are also limited by model size, also known as *largeness,* and such difficulties as *stiffness* [5] and *sparsity* [6]. Stiffness is an undesirable property of many practical Markov models (MM) and usually it is caused by: i) in case of repairable systems the rates of failure and repair differ by several orders of magnitude; ii) fault-tolerant computer systems (CS) use redundancy. The rates of simultaneous failure of redundant components are typically significantly lower than the rates

of the individual components; iii) in models of reliability of modular software the modules' failure rates are significantly lower than the rates of passing the control from a module to a module [7]. Sparsity [8] corresponds to systems, which are loosely coupled. In the subfield of numerical analysis, a sparse matrix is a matrix populated primarily with zero's [9]. If the MM is large it becomes wasteful to reserve storage for zero elements, thus solution methods that do not preserve sparsity, is unacceptable for most large problems [6].

Variety of approaches were developed to deal with MM largeness, stiffness and sparsity. They can be split into two large groups – "avoidance" and "tolerance" techniques, each contains range of advantages and limitations, thus making methods acceptable within specific conditions. In [10, 11] authors have presented the detailed description of the most common techniques, highlighting both advantages and disadvantages of their use.

Manual computations are unacceptable for large state space models. The range of specialized, off-the-shelf tools and utilities were developed to support Markov modeling process making the assessment more convenient. Still such variety or tools and techniques (T&T) can pose a difficulty when it comes to choose the most appropriate set for a specific case as every T&T is limited in its properties and applicability. The careful selection is important in case of stiffness, largeness and sparsity presence in MM, as it requires the modeler to focus on math details to avoid the use of inefficient T&T.

One of the leading standards in the safety area IEC 61508-2010 provides no special requirements for T&T, which are used to evaluate the system safety indicators, excepting the strong recommendation, that practitioner must have an understanding of the techniques used by software package to ensure its use is suitable for the specific application [2]. In contrast to the T&T, many requirements were developed for I&Cs verification and validation (V&V) tools and they are compatible by strength to the requirements of produced software and systems (see standard IEC 60880-2006 [12]). Additionally, this standard (IEC 61508-2010) asserts that methods for solving Markov models have been developed long ago and trying to improve these methods does not seem sensible. The previous works show [13, 14] that solving a large and/or stiff Markov model requires a careful selection of the solution method/tool. Otherwise, the results can differ in several orders of magnitude [14], thus, use of inappropriate method/tool for the solution of a non-trivial MM may lead to significant errors. To ensure accuracy of selected T&T for a specific case, we have applied the metric-based approach [15] during safety analysis of self-diagnostic, programmable NPP I&Cs.

In this paper we present the case-study of typical safety assessment for self-diagnostic and programmable NPP I&Cs produced by RPC "Radiy". The section 2 shows main information on I&Cs elements, structure diagram and reliability-block diagram (RBD) of analyzed architecture. Section 3 presents the developed MM. Application of metric-based approach during safety analysis and achieved results are shown in Section 4. In section 5 we present the conclusions and problems left for future research.

II. I&C ARCHITECTURE

This section presents the typical architecture of self-diagnostic and programmable NPP I&Cs produced by RPC "Radiy".

This is a Reactor Trip System (RTS) with two-channel, three-track architecture, on voting logic "2-out-of-3" for tracks in each channel and "1-out-of-2" between channels. The FPGAbased track is a basic component of observed RTS. Generally, each track can contain up to 7 module types: analogue and digital input modules (AIM, DIM); analogue and digital output modules (AOM, DOM); logic module (LM); optical communication module (OCM); and analog input for neutron flux measurement module (AIFM). The modules can be placed in 16 different positions on the track (two reserved positions for LM), using LVDS and fiber optical lines for internal/external communications.

Such flexible redundancy management helps to ensure the high availability of the system. Each channel independently receives information from sensors and other NPP systems. The channels, each being capable of forming a reactor trip signal, are independent.

The LM from each track is connected to self-diagnostic equipment, which constantly perform tests over the received data and informs on found failures and/or deviations. The diagnostic tests can only trace the found failures and have no influence on channels output data nor on general result.

In this paper, we consider the tracks consisting of five modules: LM, DIM, DOM, AIM and AOM. The Fig. 1 presents the structure diagram of a typical track. It is assumed that the corresponding components of all the tracks in the channels are identical, i.e. DIM on the 1st track is identical to the same module on other tracks in the channels, etc. The failure of the LM leads to the failure of the whole track, and failures of the DIM, DOM, AIM, AOM result in track malfunction. Therefore, it was assumed that failure of any module implies the general failed state of the track. The RDB for RTS is presented on Fig. 2. All tracks in the channels have identical hardware structures, but the software run on the system channels is diverse [16], i.e. non-identical but functionally equivalent software copies are deployed on the system channels.

Reliability index $P_{pf\,i.j}$ determines the hardware reliability of the track $T_{i.j}$ (defined by physical faults), where i indicated main ($T_{1.j}$) or diverse ($T_{2.j}$) channel, and j indicated the track number. Reliability index $P_{df\,i}$ determines software reliability of the main or diverse channels (defined by software faults), where i indicates the channel. Reliability index P_{mi}, determines reliability of majority element m_i, where $i \in (\overline{1,3})$ and D_l is the reliability of control and diagnostic tools, where $l \in \{1, 2\}$.
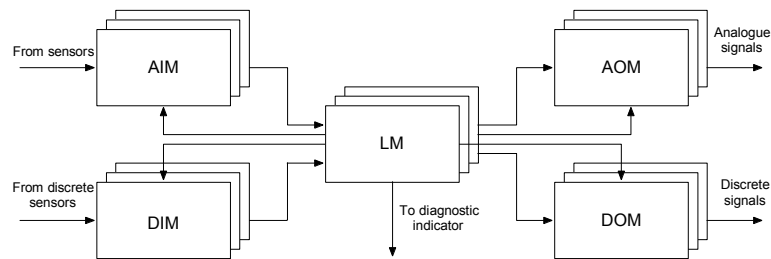

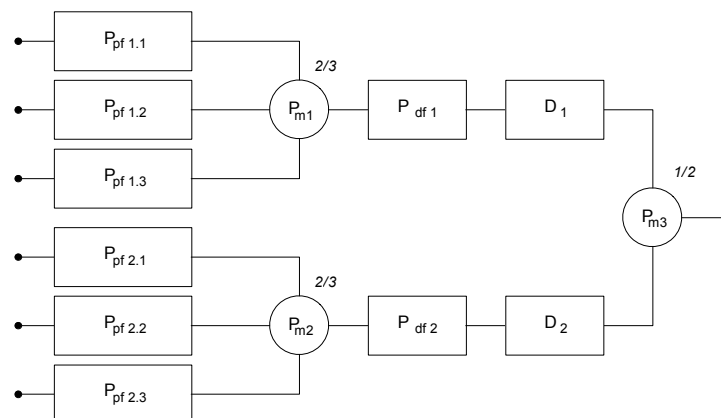Fig.1 The structure diagram of the typical track

Fig.2 The reliability-block diagram of two-channel tree-track Reactor Trip System

III. MARKOV MODEL

A.   Assumptions

The following assumptions were used to create the MM for observed RTS:
a. each element in the random moment of time can be only in two states – working and failure;
b. the system majority elements provide unstoppable correct functioning;
c. the self-diagnostic and control equipment is identical for both channels and tracks the physical faults occurrence;
d. the maintenance is performed by one group of engineers and failed chassis are repaired sequentially;
e. all detected defects are eliminated instantaneously and no new defects are introduced. The mean time between failures and mean time to repair are exponentially distributed;
f. software testing datasets are updated after each test;
g. the design faults on diverse software versions are independent events, but equal in severity. Thus, we assume that failure and repair rates for the failures caused by design faults are equal.
h. the observed RTS is FPGA-based, thus investigated software faults are such kinds of faults, which are typical for VHDL coding process that were not covered by V&V procedure. The architecture-level MM shows the rare kind of design faults that can cause a general system failure, thus we expect that not more than one undetected design fault on each software version [17, 18];
i. we assume that the rates of failure and repair of software will vary over time, e.g. as a result of executing the software in partitions as discussed in [13]. We implement this assumption based on the research work [17] that shows a plausible phenomenon – variation of software failure rates - which is well accepted in practice.

B.   Model Parameters

The MM use the following set of parameters:
a. $\lambda_{d(i)}$ – design faults failure rate, which is proportional to their residual amount n_i in i – different software versions. The change of residual amount of software faults can be present in MM using multi-fragmentation approach [17]. Using this approach, the model can be divided into N fragments that are with the same structure but may differ in one or more parameter values [17]. The number of fragments N in the MM depends on the number of expected undetected software faults N_d, the value of which can be estimated using probabilistic prediction models [9]. Based on the assumption on equivalence of design faults failure rates for diverse software the general system design failure rate is

$$\lambda_{d1} = \lambda_{d2} \rightarrow \lambda_d = \lambda_{d1} + \lambda_{d2} \ . \tag{1}$$

b. $\mu_{d(i)}$ – design failure recovery rate. Using assumption of design recovery rates equivalence, the total system design recovery rate is design recovery rate is

$$\mu_{d1} = \mu_{d2} \rightarrow \mu_d = \mu_d \Big/ \left(\sum_{i=1}^{2} \frac{\lambda_{d(i)}}{\mu_{d(i)}} \right). \tag{2}$$

c. $\lambda_{p(i,j)}$ – failure rate for the failures caused by physical faults in the track $T_{i,j}$, where $i \le 2$, $j \le 3$. As each track consist of five modules, the total $\lambda_{p(i,j)}$ of the track $T_{i,j}$ can be calculated as

$$\lambda_{p(i,j)} = \lambda_{DIM(i,j)} + \lambda_{DOM(i,j)} + \lambda_{LM(i,j)} + \lambda_{AIM(i,j)} + \lambda_{AOM(i,j)} \ , \tag{2}$$

where $\{\lambda_{DIM(i,j)}, \lambda_{DOM(i,j)}, \lambda_{LM(i,j)}, \lambda_{AIM(i,j)}, \lambda_{AOM(i,j)}\}$ – physical failure rates of DIM, DOM, LM, AIM, AOM, respectively. All corresponding components of the tracks are identical, their failure rates for the failures caused by physical faults are also equal. Thus, value $\lambda_{p(i,j)}$ is equal for all $T_{i,j}$.

d. $\mu_{p(i,j)}$ – recovery rate for the failures caused by physical faults in the track $T_{i,j}$. Equally to previous case the total recovery rate of the $T_{i,j}$ can be calculated using

$$\mu_{p(i,j)} = \lambda_{p(i,j)} / \left(\frac{\lambda_{DIM(i,j)}}{\mu_{DIM(i,j)}} + \frac{\lambda_{DOM(i,j)}}{\mu_{DOM(i,j)}} + \frac{\lambda_{LM(i,j)}}{\mu_{LM(i,j)}} + \frac{\lambda_{AIM(i,j)}}{\mu_{AIM(i,j)}} + \frac{\lambda_{AOM(i,j)}}{\mu_{AOM(i,j)}} \right) \qquad (4)$$

where $\{ \mu_{DIM(i,j)}, \mu_{DOM(i,j)}, \mu_{LM(i,j)}, \mu_{AIM(i,j)}, \mu_{AOM(i,j)}\}$ – physical failure recovery rates of DIM, DOM, LM, AIM, AOM, respectively. Using the same principle as for $\lambda_{p(i,j)}$, the $\mu_{p(i,j)}$ is equal for all system tracks.

e. D_i – the reliability of control and diagnostic tools, where $i \le 2$. Table 2 contains the MM parameters values.

TABLE I
MARKOV MODEL PARAMETERS VALUES

Parameter	λ_d	μ_d	λ_p	μ_p	D
Value	10^{-5}	0.01	10^{-4}	1	$\{0.95, 0.99\}$

The basic fragment of MM is presented on Fig. 3.

The system operates as follows. At initial moment of time all channels provide non-stop correct functioning and system is in state $S_{0(3,3)}$, where (3, 3) shows that main channel operates on 3 tracks as well as the diverse channel. At random moment the physical failure occurs on one of the tracks in main or diverse channel, and if the failure was identified by self-diagnostic tool the system moves to the state $S_{1(3,2/2,3)}$ with rate $6\lambda_p D$ in other case system moves to the state $S_{1(3,2f/2f,3)}$ with rate $6\lambda_p(1 - D)$.

The abbreviation (3,2/2,3) shows that two tracks are left in main or diverse channel and the failure was caught during diagnostic; abbreviation (3,2f/2f,3) presents situation when failure occurred in one of the channels but was not discovered by self-diagnostic tool. Such abbreviations were used to present states in basic fragment of MM.

If after the work of system maintenance group (operation in state $S_{1(3,2/2,3)}$) no new failure have occurred, the system recovers back into state $S_{0(3,3)}$ with rate μ_p. If the failure occurred on the same channel as in previous case and was identified during diagnostic the system moves to the state $S_{4(3,1/1,3)}$ with rate $2\lambda_{<}D$, and if failure was not detected – to the state $S_{5(3,1f/1f,3)}$ with rate $2\lambda_p(1 - D)$. The system recovers to the state $S_{1(3,2/2,3)}$ from $S_{4(3,1/1,3)}$ with rate μ_p. The states $S_{4(3,1/1,3)}$ and $S_{5(3,1f/1f/3)}$ presents situation when main or diverse channel goes to the failed state. In case if failure occurs in the channel, which shows constant functioning on all three tracks and was caught by self-diagnostic tools, the system moves to the state $S_{3(2,2)}$ with rate $3\lambda_p D$, in opposite case – to the state $S_{6(2f,2/2,2f)}$ with rate $3\lambda_p(1- D)$. The system recovers to the state $S_{1(3,2/2,3)}$ from $S_{3(2,2)}$ with rate μ_p.

The same process of failure occurrence, detection and elimination continuous till states $S_{11(1,1)}$, $S_{12(1f,1/1,1f)}$, $S_{13(1f/1f)}$ which shows the total system failure.

The Fig. 4 shows the transition from first MM fragment to the second based on occurrence of design failures in each state which present at least one working channel. We did not present the internal transitions in each fragment to increase readability of the design failure detection and elimination processes.
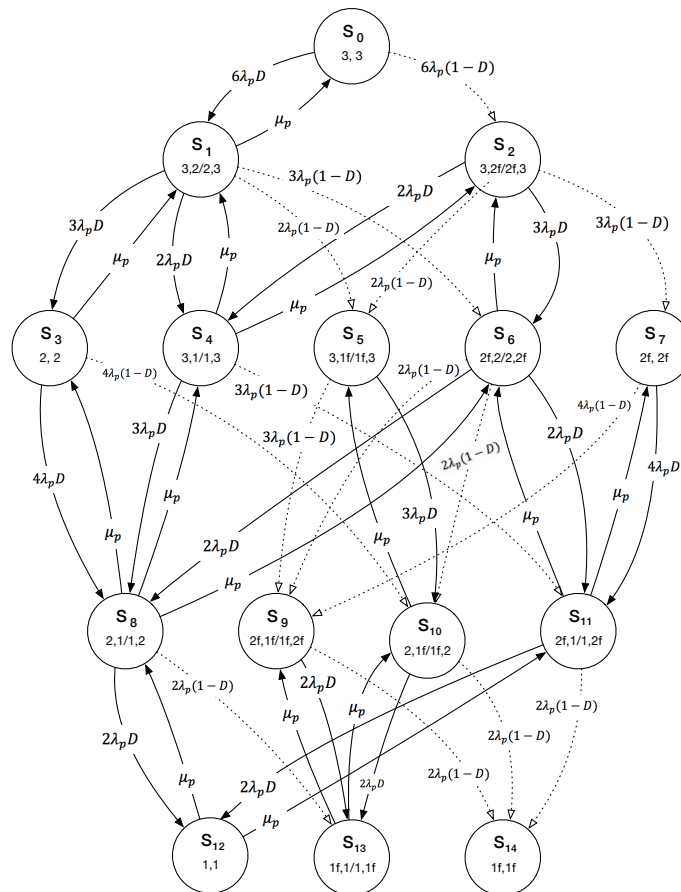
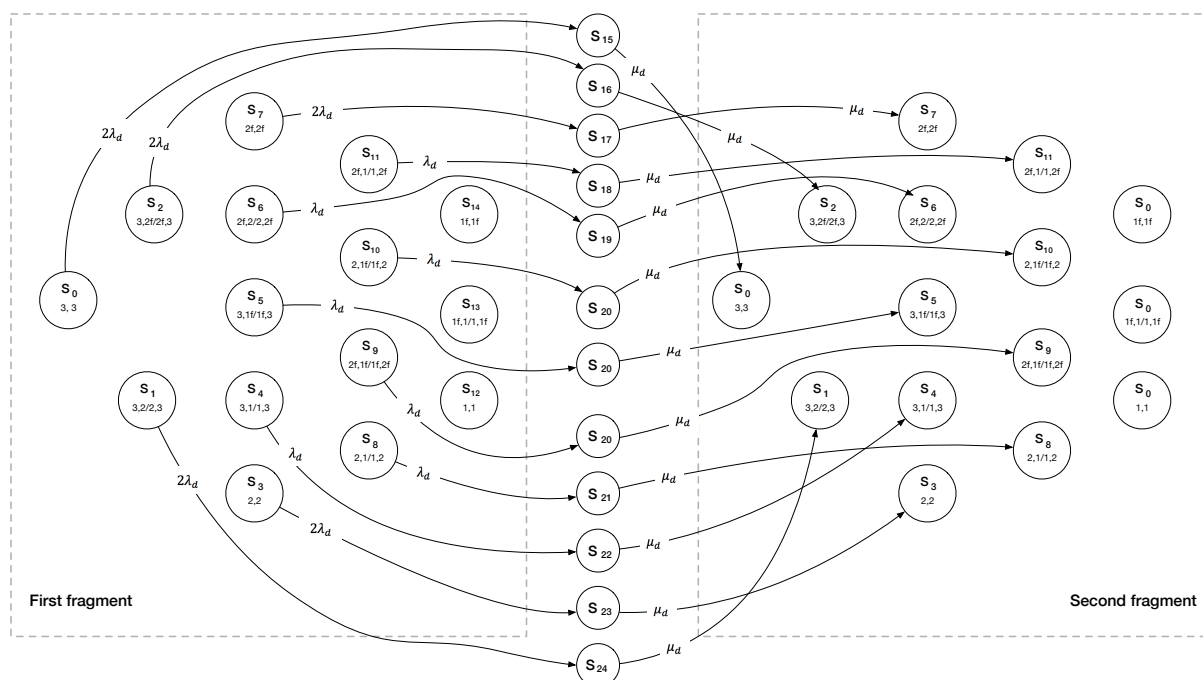Fig. 3 – Basic fragment of Markov model

Fig. 4 – Design faults detection and elimination

IV. SAFETY ANALYSIS RESULTS

We have applied the metric-based approach to select the T&T for further MM analysis. Initially it was presented in [15] and aims to reduce the T&T selection risks, increase an accuracy and optimize time and computational resources, which are spend during assessments. During the use of metric-based approach four test are consequently applied on the given MM. Each test determines on scale [0, 1] the level of four MM features, namely stiffness, decomposability, sparsity and fragmentedness, where 0 points on absence of the feature and 1 shows that given feature is strongly presented in MM. As a result, the modeler gains the recommendation of which technique is preferable for MM analysis. According to the same approach we have applied multicriteria optimization to determine the tool, which accounts following preferences: operation system support, necessary internal functions, accuracy of implemented numerical methods and use of decimal data type.

MM features test showed that presented model is moderately stiff with 0.167 value, completely decomposable with 0.02, highly sparse with 0.2 and moderately fragmented with 0.3 value. As a result, we have got the recommendation to use stiffness-tolerance methods (implicit Runge-Kutta, TR-BDF2 [11] etc.).

The multicriteria optimization was applied under the following set of preferences:
- manual (graphical or analytical) MM construction;
- functions for internal stiffness and sparsity detection;
- implemented numerical methods with accuracy at least 10^{-5}; - availability on Windows and Linux operation systems; - presence of decimal data type.

As the result we have obtained recommendation for Mathematica and/or Matlab application. To make additional verification of obtained results with selected T&T we applied utility EXPMETH, that has been validated extensively on a range of models [13 – 15, 17]. The utility uses the algorithm of modified exponential method.

To analyze the RTS safety level we can assess an availability function $A(t)$, which can be calculated as the sum of system working states probabilities, with initial condition $A(0) = 1$

$$A(t) = \sum_{i=1}^{n} P_i(t), i \in N \ , \tag{5}$$

where $P_i(t)$ is a probability of being in working state i at moment t.

The RTS working states are presented with the set $W = \{S_{0(3,3)}, S_{1(3,2/2,3)}, S_{2(3,2f/2f,3)}, S_{3(2,2)},$

$S_{4(3,1/1,3)}, S_{5(3,1f/1f,3)}, S_{6(2f,2/2,2f)}, S_{7(2f,2f)}, S_{8(2,1/1,2)}, S_{9(2f,1f/1f,2f)}, S_{10(2,1f/1f,2)}, S_{11(2f,1/1,2f)}\}$.

Fig. 5 presents the result of $A(t)$ for case $D = 0.95$ and result for D = 0.99 is shown on Fig. 6. Analysis was provided on the time interval $t \in [0; 10\ 000]$ hours.
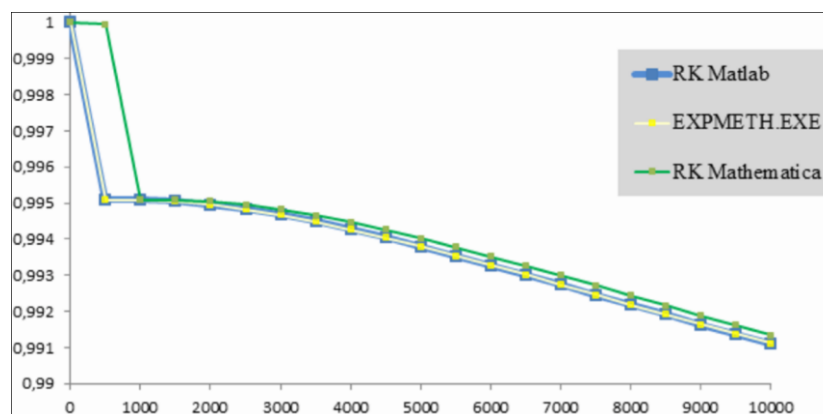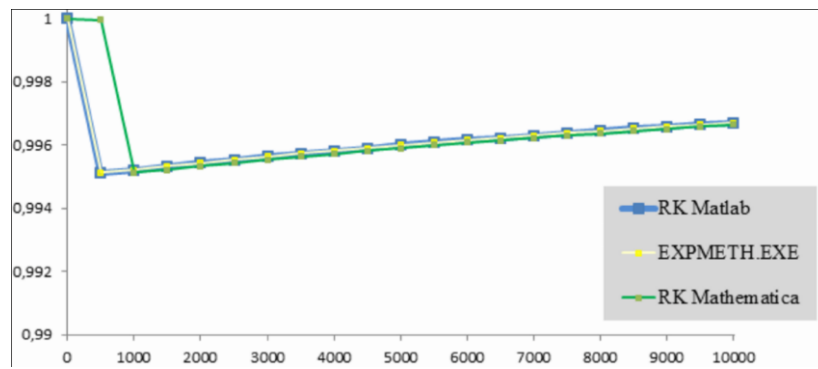

Fig. 5 RTS availability function for D = 0.95

Fig. 6 RTS availability function for D = 0.99

The minimal and maximal difference between *A(t)* values obtained with Matlab, Mathematica, EXPETH tools for D = 0.95 and D = 0.99 are presented in Table 2.

TABLE II
DIFFERENCE BETWEEN A(T) VALUES

D	Min/Max	Matlab & EXPMETH	Matlab & Mathematica	EXPMETH & Mathematica
D = 0.95	Min	$4.7*10^{-9}$	$1.3*10^{-5}$	$1.3*10^{-5}$
	Max	$1.63*10^{-6}$	$4.85*10^{-3}$	$4.86*10^{-3}$
D=0.99	Min	$1.17*10^{-8}$	$7*10_{-5}$	$7.1*10^{-5}$
	Max	$1.74*10^{-6}$	$4.82*10^{-3}$	$4.82*10^{-3}$

V. CONCLUSION

The paper describes a case study of typical RTS architecture analysis and assessment of safety parameters using the classical state-space modelling approach – Markov modelling. The metric-based approach was applied to select T&T under the set of initial preferences. Based on the obtained values we can conclude that under all assumptions presented in Section 3 and with given initial parameters values the studied architecture constructed on FPGA-based digital platform, provides the safety level which can comply the "SIL 2" according to [2].

In our future work we intend to analyze the system behavior in case if the self-diagnostic tool could not detect the design faults.

REFERENCES

[1] M. Yastrebenetskiy et al., "Nuclear Power Plants Safety: Instrumentation and control systems", Kiev: Osnova-Print, 2011, p. 768.
[2] IEC 61508, "Functional safety of electrical/electronic/programmable electronic safety-related systems", IEC Standards, 2010, p. 594.
[3] S. Archana, R. Srinivasan, K. S. Trivedi. "Availability Models in Practice", in Proc. Int. Workshop on FaultTolerant Control and Computing (FTCC-1), Seoul, 2000, pp. 1-24.
[4] ECSS-Q-ST-30-09: 2008, "European Cooperation for Space Standardization (ECSS): Availability analysis", 2008.
[5] M. Malhotra, J. K. Muppala, K. S. Trivedi, "Stiffness-Tolerant Methods for Transient Analysis of Stiff Markov Chains", Microelectronic Reliability, vol.34(11), 1994, pp.1825-1841.
[6] A. Reibman, K. S. Trivedi, "Numerical Transient Analysis of Markov models", Comput. Opns. Res., vol.15(1), 1988, pp. 19-36.
[7] A. Bobbio, K. S. Trivedi, "A Aggregation Technique for Transient Analysis of Stiff Markov Chains", IEEE Trans. on Comp., C-35, pp. 803-814, 1986.
[8] G. H. Golub, C. F. Van Loan, "Matrix Computations", JHU Press, 1996, p. 694.
[9] J. Stoer, R. Bulirsch, "Introduction to Numerical Analysis", Springer, 2002, p.732.

[10] A. Reibman, K. S. Trivedi, S. Kumar, and G. Ciardo, "Analysis of Stiff Markov Chains", ORSA Journal on Computing, vol.1(2), pp.126-133, 1989.

[11] E. Hairer, G. Wanner, "Solving Ordinary Differential Equations II: Stiff and Differential-Algebraic Problems", R. Bank, Ed. Springer, 2010, p. 631.

[12] IEC 60800, "Nuclear power plants – Instrumentation and control system important for safety – Software aspects for computer-based systems performing category A functions", 2006.

[13] V. Kharchenko, O. Odarushchenko, V. Odarushchenko and P. Popov, "Availability Assessment of Computer Systems Described by Stiff Markov Chains: Case Study", Springer, CCIS , vol. 412, 2013, pp. 112 - 135.

[14] V. Kharchenko, et. al., "Assessment of the Reactor Trip System Dependability: Two Markov's Chains – Based Cases", in Conf. Proc. Of the 10th Int. Conf. Digital Technologies, Zilina, 2014, pp. 103 – 109.

[15] V. Kharchenko et.al, "Metric-based approach and Tool for Modeling the NPP I&C System Using Markov Chains", in Conf. Proc. of 23nd ICONE, vol.5, 2014.

[16] Littlewood, B., Popov, P., Strigini, L., "Modelling Software Design Diversity - a Review. ACM Computing Surveys", vol.33 (1), 2001, pp. 177 – 208.

[17] V. Butenko, "Modeling of a Reactor Trip System Using Markov Chains: Case Study", in Conf. Proc. of 22nd ICONE, vol. 5, 2014.

[18] W. Ehrlich, et al., "Applying Reliability Measurement: A Case Study", IEEE Software, pp. 56-64, 1990.

FRIMAN

Erik Parso, Dominik Suroviak, Jaroslav Ondrejak, Marek Dragula,
Matej Jesensky, Patrik Sensel, Peter Sedlacek, Monika Vaclavkova

Abstract—The learning process of information management on Faculty of management science and informatics is quite long and hard procedure. To be a successful manager on a field of informatics that is able to lead team of developers creating some sort of information system requires to understand at least basics of programming and algorithmic thinking. Therefore, there are some subjects focusing on these abilities and skills. To make it easier to get base knowledge of programming there was founded a team to develop an application allowing students to learn faster.

Keywords—information management, unified modelling language, object oriented programming, FRIMAN, integrated development environment, flowchart diagram

I. INTRODUCTION

The purpose of information management is to:

- *design, develop, manage, and use information with insight and innovation*
- *support decision making and create value for individuals, organizations, communities, and societies [1]*

The Faculty of management Science and Informatics helps students to develop a deep understanding of the users of information and the organizational and social goals information management serves. As a result, they are equipped to use information both as a competitive tool and a means to create positive organizational change. Nowadays, these people should have a good understanding of information technologies. These technologies are used by many people to maximize efficiency of development process. Except for the basic knowledge of operation systems and office applications, these students should have also the basic know-how about the design, development and maintenance of information systems and some knowledge about the software engineering. Therefore, students studying information management at our faculty have to pass through few subjects oriented on software development. Many of them have little or none experience with programming and their way of thinking is different. It's hard for them to understand the principles of program thinking. Many software companies use an object oriented paradigm in development process. That is the main reason for students to learn to program application in Java language which is one of the most used object oriented programming language nowadays. The problem is that it is really hard for them to understand how to write a code using this language and how the program is executed. Flowchart is one of the ways to teach them the principles of algorithm thinking. But, this approach has one big disadvantage: the relation between flowchart and code is not obvious for some of them. It would be useful to provide students tool allowing them to create simple programs without having the knowledge of programming language syntax.

II. COMPARISON OF EXISTING DEVELOPMENT ENVIRONMENTS

At the beginning of the project one year ago, there was an idea to create a Java library providing simple API for application development. The purpose of this API was to make some operations like writing or reading text from console or file simpler than Java API is providing now. However, after analyses we find out that this approach is not suitable. One of our project leaders

D. Suroviak, Faculty of management science and informatics, University of Zilina, (e-mail:suroviakdominik@gmail.com)

M. Jesensky, Faculty of management science and informatics, University of Zilina, (e-mail:m.jesensky@centrum.sk)

M. Dragula, Faculty of management science and informatics, University of Zilina, (e-mail: marek.dragula@gmail.com)

J. Ondrejak, Faculty of management science and informatics, University of Zilina, (e-mail: ondrejakjaro@gmail.com)

P. Sensel, Faculty of management science and informatics, University of Zilina, (e-mail: sensel@st.fri.uniza.sk)

E. Parso, Faculty of management science and informatics, University of Zilina, (e-mail: parsoerik@gmail.com)

P. Sedlacek, Faculty of management science and informatics, University of Zilina, (e-mail: kefas555@gmail.com)

M. Vaclavkova, , Faculty of management science and informatics, University of Zilina, (e-mail: monika.vaclavkova@fri.uniza.sk)

teaches these students programming and after a few conversations with her we have identified, that these students have problem to understand program statements and how the program works. Then, we were searching for development environment that would help beginners to get better understanding of programming logic. We found many IDE's, but all of these environments expect that you have at least a basic knowledge of programming language syntax. We can now introduce some of these solutions and mention some of their advantages and disadvantages.

A. BlueJ

BlueJ is a Java development environment that is being developed and maintained at the University of Kent at Canterbury, UK, and La Trobe University, Melbourne. [2] This software is created with the purpose of introducing object-oriented programming to beginners, and its design differs from other development environments as a result. The user interface is much simpler unlike the other development environments such as NetBeans, Eclipse, etc. The environment supports interesting tools that are not available in other environments to make teaching object oriented programming faster. The best tool included in there is visualization of class structure using UML diagram. You can create the instances of objects which can be tested. You can inspect their values, invoke their methods, pass them as parameters and more. You can also directly invoke Java expressions without compiling. Thus, BlueJ is a powerful graphical shell/REPL for Java.

This IDE is widely used by our faculty in learning process. One of the reasons for choosing BlueJ was that it allows an approach where teachers and students deal with objects representing by graphical component. The pure Java language does not make learning process of object oriented programming very easy, because of numerous syntax and language details.

With BlueJ student can create an object and call its methods as the very first activity. Because users can create and interact with objects directly, concepts such as classes, objects, methods and parameters can easily be discussed in a concrete manner before looking at the first line of Java syntax. [2]

After a few years of experiences with our faculty learning process, the teachers know that this environment is not suitable for beginners, who have never met with programming so far. Program syntax, statements, input and output operations are really hard for them to understand. In conclusion, this IDE is good enough for people having at least some experience with programming. It will be much easier to learn principles of object oriented programming by using this IDE for them.
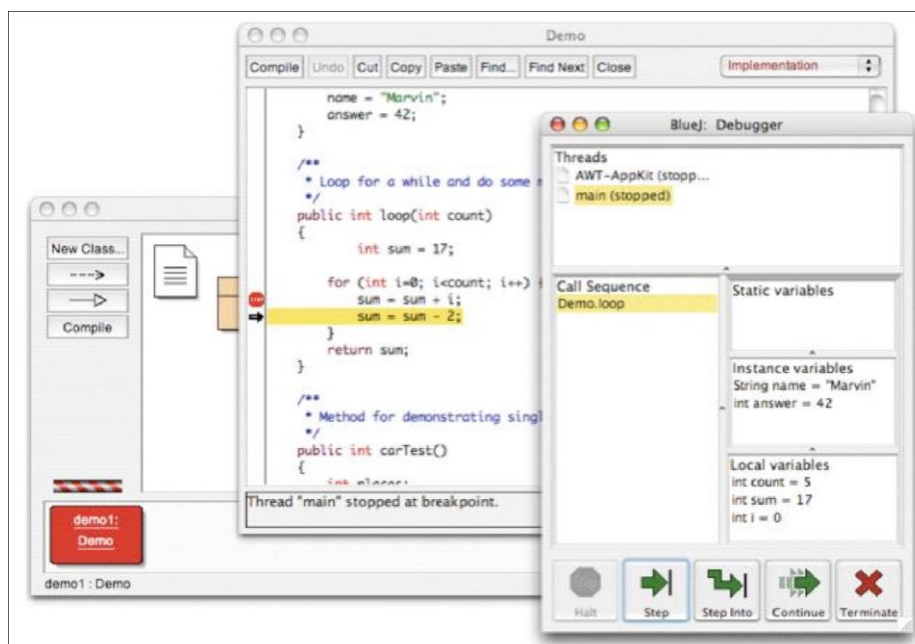
Fig 1 - BlueJ interface

Advantages:
- possibility to create instances and call methods without running whole project
- view class diagram and relationships between classes
- command interpreter
- good for better understanding of object oriented paradigm

Disadvantages:
- not suitable for big projects
- no abstraction from Java syntax
- no code generation
- intellisense is missing

B. GreenFoot

The GreenFoot is educational software designed to make learning Java programming easy and fun. This IDE is based on BlueJ. Greenfoot allows users to easily create simple graphics or visuals by providing predefined Java classes. *Greenfoot should be used to encourage the students to not give up programming if they find it difficult as it lets them immediately create the thing they were trying to do. There are many pre-defined methods such as turnLeft() or setLocation() which allow the programmer to easily move an actor around. This actor can have its controls mapped to the keyboard so that the student will be able to move his or her actor around. When the students see that they have managed to make a moving character with a few simple lines of code, they would probably be encouraged to make their Greenfoot application even more interactive. Since Greenfoot uses the actual Java code, the students are learning Java in fun way compared to the usual text based programs such as when creating a simple calculator program. [3]*

Although this environment is much simpler than BlueJ, beginners can still have problems with Java syntax. There are many environments and online solutions similar to GreenFoot. This approach of learning is better, because it is funny to create a simple game and see how objects behave when you call a method on them. The problem is, that there is no abstraction from Java syntax and you use text form of code to create program.
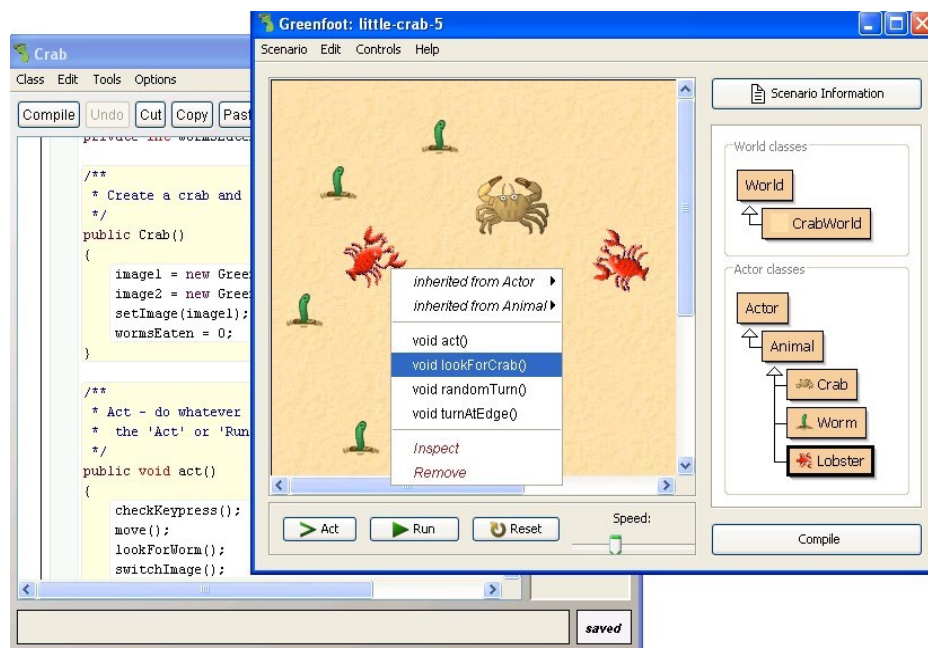
Fig 2 - Greenfoot interface

Advantages:

- possibility to create instances and call methods without running whole project
- predefined methods
- command interpreter
- visual representation of object and possibility to see the interaction with objects using methods
- predefined Java statements
- auto-completion
- syntax highlighting
- plugin system
- funny way of teaching programming

Disadvantages:
- not suitable for big projects
- no abstraction from Java syntax
- the actors are programmed in standard textual Java code
- no code generation

III. FRIMAN

There are many development environments similar to GreenFoot and BlueJ, but we have not found any that meet our expectations. And that is the main reason why we made a decision to create FRIMAN.FRIMAN will be an open source software, that has been developing within school subject since 2016. The main feature of IDE is possibility to create simple applications using flowcharts, without the knowledge of Java or any other programming language. The next image you can see shows the most important parts of this program. By using the main menu, we can create a new project, open or save existing one, run program, start debugging and many more. In the middle we can see editor. When we choose the class and method, there will be showed a flowchart diagram describing code inside of the selected method. User have a possibility to switch between Java code and flowchart representation of method. Because of it is hard for us to create diagram from Java code, code can be created just by using flowchart diagram and not by writing Java code. Diagram is created by drag and drop operations. All operations that user can do are represented by graphical item. A few of them you can see in the left down menu in picture below. These operations should provide a satisfactory support to create any functionality. In the future, we would like to provide some kind of intellisence that would make programming even easier.

At this moment, FRIMAN is divided into logic parts that communicate with each other to provide full functionality.

These parts are creating the core of FRIMAN, which consists of compiler, debugger, editor, controller, view providers, command provider etc. These parts have to communicate with each other, therefore it was necessary to design a way how these components will be connected. We have decided to use special approach how the commands will travel between application parts. Every part containing some logic is a command provider. When there is a user interaction with graphical interface, new command is created. This command is passed to controller, that will make a decision which command provider will be used to process command.

After the command is processed, the command return value will return to controller. Then controller choose the view provider that updates user interface. The same approach can be used for communication between command providers. As we mentioned before, the user of this environment does not have to have almost any knowledge of programming language syntax. This environment could be the first step, if someone wants to start programming. It will help

beginner to get better understanding of program logic. When user gets necessary knowledge how the program work, he can start to use GreenFoot or BlueJ and he can proceed to use professional environments.
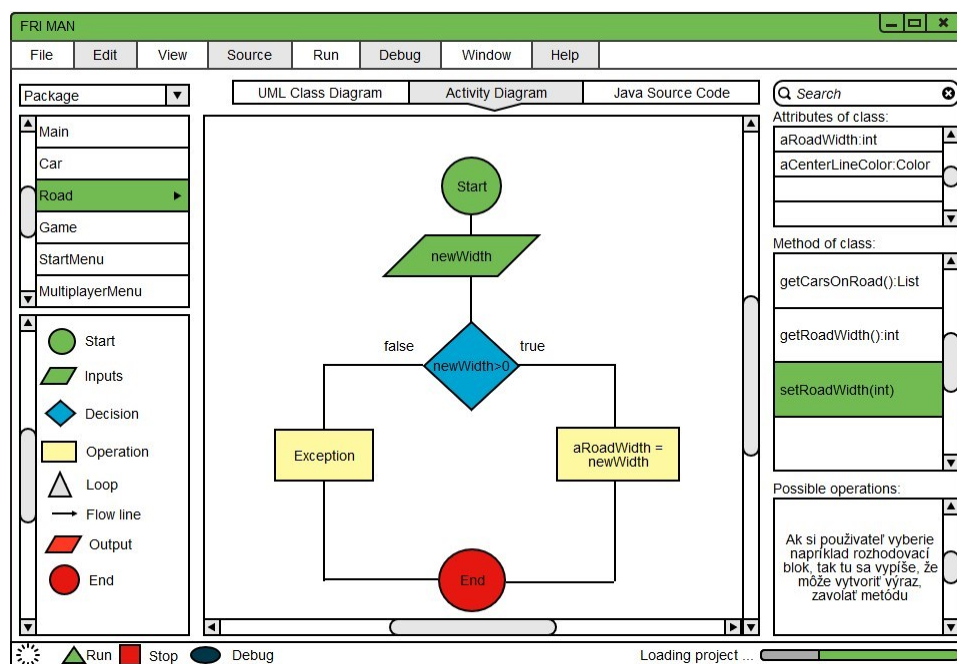

Fig. 3 - FRIMAN interface predesign

This programming IDE will provide some really useful features. Some of these features have been implemented in other environments already. One of them is the possibility to show existing classes in graphical form and to create instances of them. Then you can call methods on this instances and see the changes of these objects will be made. You can also see the return value of the method. This is the same feature that BlueJ provides right now, but we would like to make some graphical changes and make it more simple. You can see the current state of this feature in the following picture.
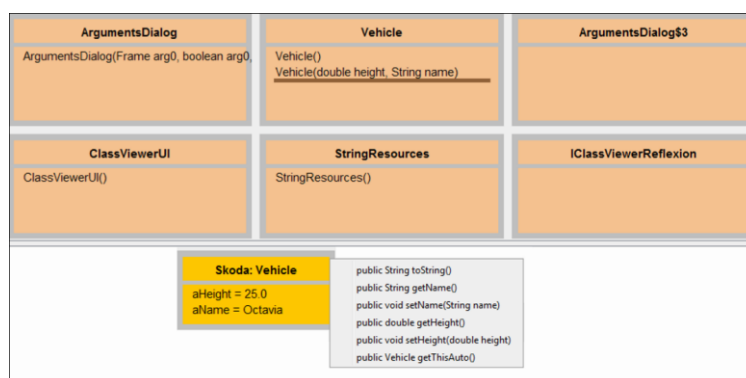

Fig. 4 - Classviewer

Moreover, there will be a code generator provided possibility to draw a flowchart and automatically create a Java code. It is one of the key parts we are working on right now and you can see its current look in image bellow.
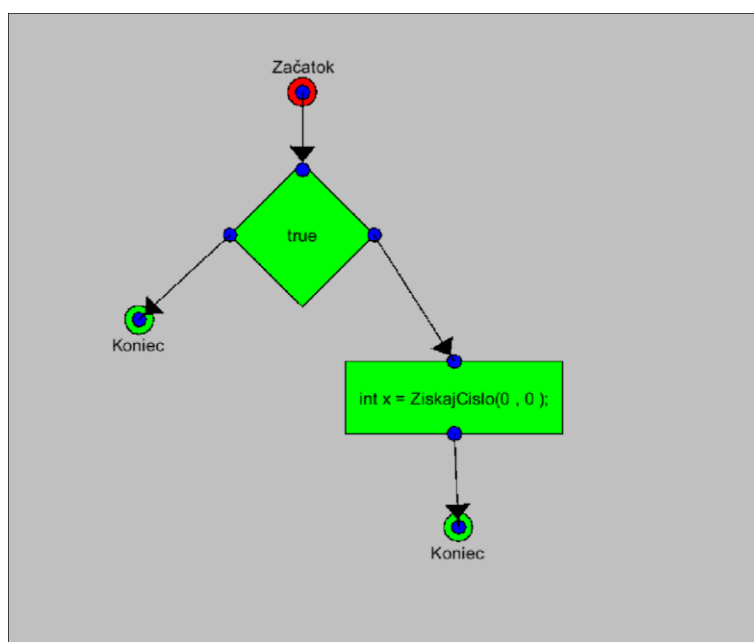
Fig. 5 - Flowchart diagram editor

As you can see at this moment this project is divided into a few parts that are developed independently but our team has already started to join them together to form a single application that allows at least some basic functionality.

IV. CONCLUSION

As mentioned before, this project started as an idea of creating a small plugin into BlueJ but after some time of problem analysis we have decided to push this idea on a whole new level. Because of its complexity, it would take much more time to complete than the time estimated at the beginning. There is a lot of things to do in the future. For example, one of the most significant parts we have not implemented yet is debugger. To make it easier and faster to work there could be some sort of intelisence with possibilities to complete half-written words, refactor code etc. If someone is interested in code optimalization there should be profiler providing information about memory consumption or method duration. In case of higher popularity there could be implemented to choose one of multiple languages.

REFERENCES

[1] "What is information management?," 2016. [Online]. Available: https://ischool.uw.edu/academics/msim/what-is-information-management. Accessed: Dec. 30, 2016.

[2] D. J. Barnes, Objects First with java: A practical introduction using BlueJ. Harlow, United Kingdom: Pearson Education, 2016.

[3] "The benefits of using Greenfoot in class," A Journey into the Teaching Experience, 2013. [Online]. Available:https://teachtolearn2202.wordpress.com/2013/10/26/the-benefits-of-using-greenfoot-in-class/. Accessed: Dec. 30, 2016.

[4] "BlueJ,". [Online]. Available: http://bluej.org. Accessed: Dec. 30, 2016.

[5] "Greenfoot,". [Online]. Available: http://greenfoot.org. Accessed: Dec. 30, 2016

Effect of boiling and microwave cooking on some antioxidant compounds in highly consumed vegetables in Egypt

Neveen F. Agamy

Abstract—Effects of boiling and microwaving methods were studied on β-carotene, Vitamin C, total phenols, antioxidant capacity, and lycopene. Vegetable samples of artichoke, green haricot, okra, pea, squash, and tomato were used for the purpose of the study. There was a statistically significant reduction in β-carotene and Vitamin C when vegetables were boiled rather than micro-waved. Total phenolic content was significantly reduced ($p<0.05$) in vegetable samples. However, a slight increase in total phenolic content of green haricot was observed when boiled (23.28%), and micro-waved (16.72%). Boiling vegetable samples resulted in a decrease in total antioxidant capacity of green haricot, squash, and tomato. On the contrary, there was an increase in total antioxidant capacity of okra, pea, and artichoke. A similar observation was made when vegetable samples were microwaved. Both decrease and increase were not statistically significant except in artichoke. There was a significant difference in flavonoids concentration ($p<0.05$) between fresh vegetables and boiled vegetables. Boiled artichoke demonstrated larger antioxidants loss than microwaved artichoke. Boiling showed a statistically significant reduction of lycopene content in tomato, while microwaving showed a non-statistically significant reduction. It can be concluded that using a microwave oven in cooking vegetables plays an essential role in maintaining higher levels of nutritive values by retaining optimum levels of most antioxidants.

I. INTRODUCTION

Eating wisely and saving food waste is a responsibility that maintains sustainability [1]. The World Health Organization (WHO) acknowledges that the global intake of vegetables is less than 20-50% of the recommended amount. In developed countries, the significantly low vegetable intake is due to the consumer's preferences for convenience foods and not the scarcity of the vegetables. Approximately 1.7 million (2.8%) of deaths worldwide are attributable to low vegetable consumption. Worldwide, insufficient intake of vegetables is estimated to cause around 14% of gastrointestinal cancer deaths, about 11% of ischemic heart disease deaths and about 9% of stroke deaths [2]. The consumption of fresh vegetables gives the consumer a variety of compounds that have a positive effect on human health. Due to the detection of many bioactive compounds in food with possible antioxidant activity, there has been an increased interest in the relationship between antioxidant and disease risks [3].

Vegetables are a good source of natural antioxidants such as carotenoids, vitamins, flavonoids, other phenolic compounds [4-7]. Epidemiological studies have shown a strong and consistent protective effect of vegetable consumption against the risk of several age-related diseases such as cancer, cardiovascular disease, cataract and macular degeneration [8-12]. To illustrate, regular consumption of tomatoes has been correlated with a reduced risk of various types of cancer [13-15] and heart diseases [16-17]. These positive effects are believed to be attributable to the high content of antioxidants, particularly lycopene [16]. Further, carotenoids are another type of antioxidants that have beneficial effects well-documented in the literature. In addition to their role as a precursor of vitamin A, carotenoids are valuable antioxidants, help in the prevention of atherosclerosis [18]. For example, β -Carotene is the well-known pro vitamin A, which can play a valuable role in promoting visual health. On the other hand, a wide

N.F.Agamy, High Institute of Public Health, Alexandria University, Egypt (e-mail: nevagamy@gmail.com).

range of non-vitamin A active carotenoids, have been shown to be primary components of the human macula pigment [19].

Nowadays, there is a global tendency to improve sustainable consumption patterns in both developed and developing countries through cooking devices such as microwave ovens. In Egypt, boiling is the most conventional method in cooking vegetables, whereas microwaving is a method that has been introduced only recently. These cooking methods are estimated to bring about a number of changes in physical characteristics and chemical composition of vegetables [12, 20], for example, [21] showed that boiling and baking had a small effect on the ascorbic acid, total phenolic, lycopene and antioxidant activity of the tomatoes. [12] indicate that cooking affected the antioxidant components and antioxidant activity of some vegetables. [22] found that thermal treatment decreased the total phenolic content in all vegetables, and antioxidant activity in some of them. [23] pointed out that processed vegetables show a wide range of phytochemical loss, and the technology in the food industry should be used to reduce the loss of antioxidants and micronutrients to the minimum by means of mild processes and the monitoring of each step of the transformation with due control assays. Therefore, the current study compares the effect of two cooking methods: boiling versus microwaving, on the antioxidants contents of vegetables; it aims at investigating best practices to maintain optimum nutritive value when boiling or microwaving vegetables. The study determines the concentrations of β carotene, vitamin C, total phenols, total antioxidant capacity, flavonoids, and lycopene in the following vegetables: artichoke, green haricot, okra, pea, squash, and tomato.

II. Materials and Methods

Sampling. Fresh vegetables are brought daily from villages to local markets in Alexandria in Egypt. Three kilograms of each fresh artichoke, green haricot, okra, pea, squash, and tomato were purchased in October 2013 at a local market in Alexandria, to be used as research materials. The samples were randomly selected off the shelf.

Preparation of Vegetable Samples. Vegetables were washed with tap water after removing inedible parts with a sharp knife. Vegetables dried on paper towel and were cut into small pieces or slices almost equal in size, then mixed well by hand to get a representative batch. Nine hundred grams were taken and divided into three portions (300 g for each application). One portion was retained raw; the other two portions were cooked using two different methods in triplicate, as explained below.

Boiling. One hundred and fifty ml of tap water was heated to reach the boiling point in a stainless steel pan. An amount of 100 grams of vegetable samples was added to the boiling water to be cooked for 5 minutes. The samples were drained off and cooled rapidly.

Microwave Cooking. An amount of 100 grams of vegetable samples was put in a glass dish. Six ml of tap water was added to the samples. The dish was covered with a cooking bag that had several holes. The samples were cooked for 5 minutes in a commercial Microwave Panasonic oven NN-S 2/5 WF, 800 watt/energy, 22L/capacity, 220 volt. The samples were drained off and cooled rapidly on ice.

Analytical Methods. Raw and cooked vegetables were homogenized in a blender (Moulinex–France) for 2 minutes for further analysis.

Determination of moisture content. It was carried out according to [24], the chemical analysis of fresh vegetables or just cooked was expressed on wet weight basis [25].

Determination of Vitamins and antioxidant content:

     A. Determination of β carotene by HPLC

It was determined according to the method of [26]. For carotenoids extraction,0.5 g of

homogenized vegetable sample was extracted in 500 ml acetone: ethyl acetate (2:1, v/v)for 1 h in the dark. The extraction mixture was centrifuged at 10,000 rpm for 15 min. The organic phase was transferred to a new micro centrifuge tube and was evaporated under vacuum. The residues were dissolved in 100 ml ethyl acetate and 5 ml was injected on reverse phase HPLC. Analyses of carotenoids and ascorbic acid were performed using an Agilent 1200 series HPLC system (Santa Clara, CA), including a model G1311A quaternary pump, a model G1367Bauto sampler, a model G1316A column oven, and a model G13150 photodiode array detector. The column used was an Agilent ZORBAX Eclipse XDB-C18, 5 mm bead size, 4.6 mm_150 mm, connected with an Eclipse XDB-C18 guard column. The column temperature was controlled at 308°C during the HPLC runs. The flow rate was at 1 ml/ min. The mobile phases were acetonitrile:H2O: tri ethylamine (900:99:1,v/v/v) (A) and ethyl acetate (B). The gradient elution program was:0–5 min, 100–75% A; 5–10 min, 75–30% A ; 10–13 min, 30–0% A; 13–14 min,0–100% A; 14–15 min, 100% A. Data were collected at 440 nm, 477 nm and 296 nm. β-Carotene was identified based on the retention time and the spectrum as compared to the commercially available authentic standards.

B. Determination of Vitamin C by HPLC

Vitamin C was extracted according to the modified method of [27].The sample (10 g) was homogenized with an extracting solution containing meta-phosphoric acid (0.3M) and acetic acid (1.4 M). The mixture was placed in a conical flask (wrapped with aluminum foil) and agitated at 100 rpm with the aid of an orbital shaker for 15 min at room temperature. The mixture was then filtered through a Whatman No. 4 filter paper to obtain a clear extract. The ratio of the sample to extraction solution was 1 to 1. All samples were extracted in triplicates.

C. Determination of total phenolic and total antioxidant determination

The extraction procedure was a modification of the method described by [28]. Edible portion of each wet plant material was homogenized using blender immediately before extractions. A quantity (50 g) of each ground plant material was weighed separately, and 250 ml of50% aqueous ethanol (1:5 w/v) was added and mixed in vertical shaker for 6 h at 40°C inconstant temperature bath which did not vary more than two degrees either way. Then, the liquid extract was filtered and centrifuged at 2000 rpm for 20 min at room temperature to obtain a clear supernatant liquid which was used directly for DPPH and total phenolic compounds.

The antioxidant activity of the extracts, on the basis of the scavenging activity of The 2,2-Di (4-tert-octylphenyl)-1-picrylhydrazyl (DPPH) free radical, was determined by the method described by [29], 0.5 ml solution of DPPH radical in methanol was freshly prepared daily. An aliquot of 15-140 µl of each extract solution was mixed with 100 µl of methanolic DPPH radical to give final concentrations of 3- 28 mg extract/ml DPPH. The percent inhibition of radical scavenging ability was calculated as follow; % inhibition = [(Absorption of standard solution of ascorbic acid - Absorption of samples)/ Absorption of standard solution of ascorbic acid] *100

The amount of total phenolic contents in the extracts was determined calorimetrically with the Folin-Ciocalteu (FC) reagent, 20 µl of extract and 100 µl of undiluted (FC) reagent were added with incubation at 40°C in the dark for 30 min for colour development was performed and then reading absorbance at 765 nm by UV-1650PC visible spectrophotometer. The absorbance obtained was converted to gallic acid equivalent as milligrams per gram (GAE/g), using gallic acid standard curve.

D. Determination of Flavonoids (hesperidin) by HPLC

Hydrolysis, extraction, and recovery test according to [30], where100 mg sample placed in a 20 ml tube containing 10 mg ascorbic acid dissolved in 5 ml of acidified methanol (1.2 M HCl)

was flushed with N2 air for 30 sec and then refluxed at 80°C for 2hour. After cooled down to room temperature, the sample was centrifuged at 4000g for 10 min. Supernatant, approximately 2 ml was taken and filtered through 0.2 μm syringe filter (Millipore, Bedford, MA). The filtrate was kept at 10°C for HPLC analyses within 12 h. Flavonoids (hesperidin) were separated using the HPLC system equipped with a Water 2695 separation module and an Agilent Zorbax ODS column (3.5μm, 4.6 x 150 mm) at 35 °C using a gradient from0 −15 min, 1 to 25% acetonitrile (ACN) in 1% aqueous formic acid (FA); and 15−50 min, 25%−40% ACN in 1% aqueous FA at a flow rate of 0.7 ml/min. The column elute was monitored using a Waters 2996 photo diode array detector (250−700 nm). Identification and quantification of individual flavonoids was carried out using commercial standards.

E. Determination of lycopene

Lycopene concentration of tomato was measured using a HPLC system described by [21, 31].

Statistical analysis. All data were recorded as means ± SE and Duncan comparisons were carried out to test any significant differences between raw and cooked vegetables. The quantitative data presented for each treated vegetable was based on the average of three replicates and expressed on a wet weight basis (n= 3).

Chemicals. DPPH and Folin–Ciacalteu reagents were purchased from Sigma Aldrich Egypt, Other chemicals used were all analytical grade and bought from "Al Gomhoreya" company-Egypt.

III. RESULTS

Results in Tables I-VI represent the concentrations of β -Carotene, ascorbic acid, total phenolic compounds, lycopene and antioxidant activity of the cooked vegetables compared to raw.

TABLE I:
EFFECT OF BOILING AND MICROWAVE COOKING PRACTICES ON B CAROTENE CONTENT (MG/100G)

Vegetable	Fresh Mean ± SD	Boiled Mean ± SD	% Change	Microwave Mean ± SD	% Change
Artichoke	1.37 ± 0.19	0.27 ± 0.03	19.59	1.26 ± 0.14	8.04
Green haricot	1.42 ± 0.16	0.43 ± 0.05**	69.50	1.27 ± 0.14	11.10
Okra	1.49 ± 0.17	0.87 ± 0.10**	42.22	1.30 ± 0.14	13.14
Pea	2.24 ± 0.25	0.86 ± 0.10**	61.72	1.61 ± 0.18	28.16
Squash	2.14 ± 0.24	0.62 ± 0.07**	71.02	1.53 ± 0.17	28.60
Tomato	5.96 ± 0.66	1.29 ± 0.14**	78.22	3.20 ± 0.24	46.32

Data are expressed as means ± SE of triplicate experiments (on wet basis)*Significant at p < 0.05. **Significant p < 0.01

TABLE II.
EFFECT OF BOILING AND MICROWAVE COOKING PRACTICES ON VITAMIN C CONTENT (MG/100G)

Vegetable	Fresh Mean ± SD	Boiled Mean ± SD	% Change	Microwave Mean ± SD	% Change
Artichoke	7.40 ± 0.82	5.20 ± 0.69	29.72	6.20 ± 0.58	16.21
Green haricot	7.90 ± 0.88	1.79 ± 0.20*	77.34	5.60 ± 0.62	29.11
Okra	6.10 ± 0.68	3.50 ± 0.39*	42.62	5.40 ± 0.60	11.47
Pea	3.79 ± 0.42	2.10 ± 0.23	44.59	3.30 ± 0.37	12.92
Squash	11.20 ± 1.24	1.60 ± 0.18**	85.71	10.6 ± 1.18	5.35
Tomato	12.79 ± 1.42	3.50 ± 0.39*	72.63	7.45 ± 0.02	41.75

Data are expressed as means ± SE of triplicate experiments (on wet basis)*Significant at p<0.05.**Significant p<0.01

TABLE III.
EFFECT OF BOILING AND MICROWAVE COOKING PRACTICES ON TOTAL PHENOLIC CONTENT OF DIFFERENT VEGETABLES (MG/100G)

Vegetable	Fresh Mean ± SD	Boiled Mean ± SD	% Change	Microwave Mean ± SD	% Change
Artichoke	0.11 ± 0.21	0.054± 0.01*	49.05	0.09 ± 0.02	15.09
Green haricot	0.34 ± 0.03	0.413 ±0.18	+23.28	0.39 ± 0.02	+16.72
Okra	0.84 ± 0.26	0.02 ± 0.01**	72.599	0.53± 0.05*	37.57
Pea	0.11 ± 0.33	0.05 ± 0.01*	52.22	0.09 ±0.02	20.37
Squash	0.09 ± 0.36	0.04 ± 0.09**	63.13	0.08 ±0.02	24.24
Tomato	2.18 ± 0.31	1.41 ± 0.12*	35.32	1.82 ± 0.01	16.51

Data are expressed as means ± SE of triplicate experiments (on wet basis)*Significant at p<0.05. **Significant p<0.01

TABLE IV.
EFFECT OF BOILING AND MICROWAVE COOKING PRACTICES ON TOTAL ANTIOXIDANT CAPACITY OF VEGETABLES

Vegetable	Fresh Mean ± SD	Boiled Mean ± SD	% Change	Microwave Mean ± SD	% Change
Artichoke	46.74 + 0.03	68.60 + 0.02*	+ 46.77	69.53 + 0.01*	+48.76
Green haricot	63.02 + 0.03	60.69 + 0.02	3.70	44.69 +0.01	29.01
Okra	65.58 + 0.02	67.81 + 0.02	+3.40	85.11 + 0.01*	+29.78
Pea	47.67 +0.03	51.02 + 0.02	+7.03	43.58 + 0.04	8.58
Squash	71.16 + 0.02	53.72 + 0.01	24.51	42.25 + 0.02*	40.63
Tomato	57.40 + 0.02	66.50 + 0.02	15.85	59.23 + 0.02	+3.19

Data are expressed as means ± SE of triplicate experiments (on wet basis)*Significant at p<0.05.

TABLE V.
HESPERIDIN'S CONTENT OF VEGETABLES AT BOILING AND MICROWAVE COOKING (MG/100G)

Vegetable	Fresh Mean ± SD	Boiled Mean ± SD	% Change	Microwave Mean ± SD	% Change
Artichoke	94.00 ± 10.44	2.05 ± 0.13**	97.81	24.60 ± 2.73*	73.82
Green haricot	2.73 ± 0.30	0.96 ± 0.02*	64.83	1.86 ± 0.62	31.86
Okra	76.00 ± 8.44	2.04 ± 0.23**	74.31	24.08 ± 2.68*	68.31
Pea	78.61 ± 8.73	5.10 ± 0.57**	93.51	69.38 ± 0.68*	11.74
Squash	5.18 ± 0.58	1.17 ± 0.13*	77.41	2.60 ± 0.18	49.80
Tomato	10.19 ±1.13	6.57 ±0.42*	35.52	8.87 ±10.67	12.95

Data are expressed as means ± SE of triplicate experiments (on wet basis)*Significant at p<0.05. **Significant p<0.01

TABLE VI.
LYCOPENE CONTENT IN TOMATO AT DIFFERENT COOKING TREATMENTS (MG/100G)

Vegetable	Fresh Mean ± SD	Boiled Mean ± SD	% Change	Microwave Mean ± SD	% Change
Tomato	7.80 ± 0.34	3.20 ± 0.36*	59.03	5.90 ± 0.83	24.35

Data are expressed as means ± SE of triplicate experiments (on wet basis)*Significant at p<0.05

IV. DISCUSSION

A. Effect of boiling and microwave cooking practices on β carotene

Table I compares the concentration of β-carotene in fresh vegetables and its concentration in both boiled and micro-waved vegetables. The highest loss of β -carotene (80%) was detected when artichoke was boiled. In contrast, when microwaving artichoke, only 8.04% of B-carotene was lost. There was a significant difference in β -carotene concentration (p<0.05) between fresh vegetables and boiled vegetables. On the other hand, no statistically significant difference was detected in β -carotene between fresh vegetables and micro-waved vegetables. Both boiling and microwaving caused a substantial loss of β-carotene, which could be explained by the consequent leaching of molecules into water and their instability at the boiling temperatures of the boiling process (100° C). Some researchers have reported similar results of β carotene loss from vegetables, including spinach, amaranth and fenugreek, during cooking procedures, such as boiling, stewing, frying, blanching and pressure cooking [32, 33]. In contrast, [25] found no lutein content in the water after boiling vegetables in it, suggesting that no carotenoids leached out water during the boiling of the vegetables.

B. Effect of boiling and microwave cooking practices on ascorbic acid (vitamin C) content

Table II compares the concentration of ascorbic acid (Vitamin C) in fresh vegetables and its concentration in both boiled and micro-waved vegetables. To start with, the highest loss of Vitamin C (85.71%) was detected when squash was boiled causing highly statistical significance (p<0.001). In contrast, when microwaving squash, only 5.35% of Vitamin C was lost. There were other significant differences in Vitamin C concentrations between fresh and boiled green haricot, okra, and tomato (p<0.05). Artichoke and pea showed no significant difference in Vitamin C between fresh and boiled samples. No statistically significant difference was detected in Vitamin C between fresh vegetables and micro-waved vegetables. This indicates that cooking affects the retention of ascorbic acid in the tissues. [34] pointed out that the cooking procedures could result in significant losses of vitamin [32, 33] reported losses

of ascorbic acid from vegetables including spinach and fenugreek, during cooking procedures, such as boiling, stewing, frying, blanching, and pressure cooking.

C. Effect of boiling and microwave cooking practices on total phenolic acid

Table III compares the concentration of total phenols in fresh vegetables and its concentration in both boiled and micro-waved vegetables. Total phenolic content was significantly reduced (p<0.05) in vegetable samples. However, a slight increase in total phenolic content of green haricot was observed when boiled (23.28%), and micro-waved (16.72%). This was not statistically significant. Research data on total phenols in cooked vegetables is very limited. [12] reported that raw broccoli floret retained (28.1%) and (28.4%) of total phenolic content when boiled and micro-waved respectively. [35] reported that boiling broccoli for 15 minutes retained 18% of total phenols. Remaining total phenols leached into cooking water. [36] found that after blanching, total phenols decreased or increased depending on the type of vegetables, a result which is in consistency with the current study.

D. Effect of boiling and microwaving on the total antioxidant capacity (mg/g GAE)

Table IV compares the concentration of total antioxidant capacity (mg/g GAE) in fresh vegetables and its concentration in both boiled and micro-waved vegetables. Boiling vegetable sample resulted in a decrease in total antioxidant capacity of green haricot, squash, and tomato. On the contrary, there was an increase in total antioxidant capacity of okra, pea, and artichoke. Both decrease and increase were not statistically significant except in artichoke. To illustrate, boiling and microwaving artichoke caused the highest significant increase in total antioxidant capacity (46.77%, 48.76% respectively). The study by [12] showed that there was no significant difference in total antioxidant capacity between boiling and microwaving. Peas had a lower total antioxidant capacity than spinach [22], which conforms to the results of the present study. [12] reported that raw broccoli florets had 60.5% of total antioxidant capacity as measured by DPPH when boiled and micro-waved; the florets retained 35% and 34.7% of total antioxidant capacity respectively. However, results of the current study showed that total antioxidant capacity of raw vegetables was less than that reported by [12]. [22] reported that total antioxidant capacity of vegetables boiled for one minute was similar to the capacity of fresh ones.

E. Effect of boiling and microwave cooking practices on Flavonoids (hesperidins) content

Table V compares the concentration of flavonoids in fresh vegetables and their concentration in both boiled and micro-waved vegetables. The highest loss of flavonoids (97.81%) was detected when artichoke was boiled. Similarly, when microwaving artichoke, 73.82% of flavonoids were lost. There was a significant difference in flavonoids concentration (p<0.05) between fresh vegetables and boiled vegetables. On the other hand, no statistically significant difference was detected in flavonoids between fresh vegetables and micro-waved vegetables except in artichoke (73.82%) and pea (68.31%). Both boiling and microwaving caused a substantial loss of flavonoids, which could be explained by the consequent leaching of molecules into water and their instability at the boiling temperatures of the boiling process (100° C).

Flavonoids are potent antioxidant compounds found in plants that have been found, for instance, to inhibit tumour development [37]. Flavonoids also have a wide range of other potential benefits [38, 39].

F. Effect of boiling and microwaving on lycopene content in tomato

Table VI illustrates the concentration of lycopene in fresh tomato and its concentration in both boiled and micro-waved tomato. Boiling tomato decreased lycopene content by 59.03%,

which was statistically significant. Likewise, when tomato was micro-waved, lycopene content was reduced by 24.35% showing no statistical significance. Similar studies reported that lycopene in tomato was dramatically dropped during microwave heating and baking [40, 41].

V. Conclusion

To begin with, there was a statistically significant reduction in β-carotene and Vitamin C when vegetables were boiled, whereas there was no significant difference when micro-waved. Consequently, it can be concluded that micro-waved vegetables retain higher amount of β-carotene and Vitamin C than boiled vegetables. Further, total phenolic content was significantly reduced in vegetable samples. However, a slight increase in total phenolic content was detected in both boiled and micro-waved green haricot, which was not statistically significant. Moreover, boiling and microwaving almost had the same effect on total antioxidant capacity. However, levels of total antioxidant capacity vary according to the type of vegetable. In addition, both boiling and microwaving caused a substantial loss of flavonoids. This loss was statistically significant in boiling, whereas insignificant in microwaving vegetables. Likewise, boiling showed a statistically significant reduction of lycopene content in tomato, while microwaving showed a non-statistically significant reduction. It can be concluded that using a microwave oven in cooking vegetables plays an essential role in maintaining higher levels of nutritive values by retaining optimum levels of most antioxidants.

VI. Recommendations

It is highly recommended to use microwave devices to help towards better retention of antioxidants in vegetables. Moreover, as there is a loss of antioxidant compounds in boiled vegetables, shorter cooking time guarantees better antioxidant retention. Furthermore, the person in charge of preparing food is advised to consider food nutritional quality by choosing cooking methods that avoid the loss of antioxidants. In addition, in this study, it is estimated that using household devices (microwave oven) could save natural antioxidants and avoid depending on antioxidant supplementation. Nevertheless, the results of this study should be used in organizing workshops and seminars to enhance health education. This should take place on a narrow scope represented in family, as well as a broader scope including catering in hotels, restaurants, hospitals, etc. Last but not least, further studies are needed to investigate the effect of cooking times and conventional Egyptian cooking methods on the nutritional quality of food.

References

[1] UNESCO 2013 Message from Ms Irina Bokova, Director-General of UNESCO, on the occasion of World Environment Day, 5 June 2013. http://www.unesco.org/new/en/tashkent/aboutthisoffice/singleview/news/ message_from_ms_irina_bokova_director_general_of_unesco_on_the_occasion_of_world_environment_da y_5_june_2013/#.UlMf_lBHL44 accessed on 6.10.2013

[2] WHO 2013 "Global Strategy on Diet, Physical Activity and Health. Promoting fruit and vegetable consumption around the world." Information sheet. http://www.who.int/dietphysicalactivity/fruit/en/ index2.html

[3] J.Nilsson, R.Stegmark, B.Akesson, Total antioxidant capacity in different pea (Pisumsativum) varieties after blanching and freezing, *Food Chem,* vol. 86, 2004, pp.501–507.

[4] J.H.Cohen, A.R.Kristal, J.L.Stanford, Fruit and vegetable intakes and prostate cancer risk. *Journal of National Cancer Institute* , vol. 92, 2000, pp. 61–68.

[5] S.Liu, J.E. Manson, I. M.Lee, S. R.Cole, C. H.Hennekens, etc., .Fruit and vegetable intake and risk of cardiovascular disease, *The Women_s Health Study.American Journal of Clinical Nutrition*, vol. 72, .2000, pp.922–928.

[6] M.I.Sweeney, W.Kalt, S. L.Mackinnon, J.Ashby, K.T.Gottschall- Pass, Feeding rats diets enriched in low bush blueberries for six weeks decreases ischemia-induced brain damage. *Nutritional Neuroscience*, vol. 5, no.6, 2002, pp.427–431.

[7] Y.S.Velioglu, G.Mazza, L.Gao, B.D.Oomah, Antioxidant activity and total phenolic sin selected fruits, vegetables, and grain products, *Journal of Agricultural and Food Chemistry*, vol.46, 1998, pp.4113–4117.

[8]   L.M.Cheung, P.C.K.Cheung, V.E.C.Ooi, Antioxidant activity and total phenolics of edible mushroom extracts, *Food Chem*, vol. 81, 2003, pp.249–255.

[9]   K.J.Hunter, J.M.Fletcher, The antioxidant activity and composition of fresh, frozen, jarred and canned vegetables, *Innovative Food Science and Emerging Technologies*, vol. 3, 2002, pp. 399–406.

[10] K.E.Heim, A.R.Tagliaferro, D.J.Bobilya, Flavonoid antioxidant: chemistry, metabolism and structure–activity relationships, *Journal of Nutritional Biochemistry*, vol. 13, 2002, pp. 572–584.

[11] W.Lopaczynsk, S.H.Zeisel, Antioxidants, programmed cell death, and cancer, *Nutrition Research*., vol. 21, 2001, pp. 295–307.

[12] D.Zhang, Y.Hamauzu, Phenolics, ascorbic acid, carotenoids and antioxidant activity of broccoli and their changes during conventional and microwave cooking, *Food Chemistry*, vol.88, 2004, pp.503–509.

[13] S.Franceschi, E.Bidoli, C.La Vecchia, R.Talamini, B.D'Avanzo, E.Negri, Tomatoes and risk of digestive tract cancers, *International Journal of Cancer*, vol. 59, 1994, pp.181–184.

[14] H.Gerster, The potential role of lycopene in human health, *Journal of American College of Nutrition*, vol. 16, 1997, pp. 109–126.

[15] J.H.Weisburger, Evaluation of the evidence on the role of tomato products in disease prevention, *Proc. of the Society of Experimental Biology and Medicine*, vol. 218, 1998, pp. 140–143.

[16] V.Lavelli, C.Peri, A.Rizzolo, Antioxidant activity of tomato products as studied by model reactions using xanthine oxidase, my eloper oxidase, and copper-induced lipid peroxidation, *Journal of Agricultural and Food Chemistry*, vol. 48, 2000, pp.1442–1448.

[17] D. K.Pandey, R.Shekelle, B.J.Selwyn, C.Tangney, J.Stamler, Dietary vitamin C and beta carotene and risk of death in middle-aged men, *American Journal of Epidemiology*, vol. 142, 1995, pp. 1269–1278.

[18] S. M.Moeller, P. F.Jacques, J. B.Blumberg, The potential of dietary xanthophylls in cataract and age-related macular degeneration, *Journal of the American College of Nutrition*, vol. 19, no. 5, 2000, pp. 522-527.

[19] R.A.Bone, J.T.Landrum, S.L.Tarsis, Preliminary identification of the human macularpigment, *Vision Research*, vol. 25, no. 11, 1985, pp. 1531-1535.

[20] Z. U.Rehman, M.Islam, W. H.Shah, Effect of microwave and conventional cooking on insoluble dietary fibre components of vegetables, *Food Chemistry*, vol. 80, 2003, pp.237–240.

[21] E.Sahlin, G. P.Savage, C. E.Lister, Investigation of the antioxidant properties of tomatoes after processing, *Journal of Food Composition and Analysis*, vol. 17, 2004, pp.635-647.

[22] A.Ismail, Z.M.Marjan, C.W.Foong, Total antioxidant activity and phenolic content in selected vegetables, *Food Chem*, vol. 87, 2004, pp.581–586.

[23] M.Blasa, G.Lorenzo, A.Donato, N.Paolino, Fruit and Vegetable. Antioxidants in Health. Chap.3, *Bioactive Foods in Promoting Health*. Ed. R.Watson, V. Preedy, 2010, pp.37–58.

[24] AOAC. Official Methods of Analysis of AOAC International, AOAC, vol. 16, 1995, pp.816-851.

[25] Y.Liu, C. O.Perera, V.Suresh, Comparison of three chosen vegetables with others from South East Asia for their lutein and zeaxanthin content, *Food Chem*., vol. 101, no.4, 2007, pp.1533-1539.

[26] H.S.Lee, W.S.Castle, G.A.Coates, High performance liquid chromatography for the characterization of carotenoids in the new sweet orange (early gold), *U.S.A J. Chromatogr*., vol. 913, 2001, pp. 371-377.

[27] A.Abdulnabi, A.H.Emhemed, G.D.Hussein, Determination of antioxidant vitamin sin tomatoes, *Food Chem*., vol. 60, 1997, pp. 207-212.

[28] B.Lapornik, M.Prosek, A.G.Wondra, Comparison of extracts prepared from plant by-products using different solvents and extraction time, *Journal of Food Engineering*, vol. 71, no. 2, 2005, pp.214-222.

[29] L.P.Leong, G.Shui, An investigation of antioxidant capacity of fruits in Singapore markets, *Food Chem*., vol.76, 2002, pp.69-75.

[30] K.H.Miean, S.Mohamed, Flavonoid (myricetin, quercetin, kaempferol, luteolin, and apigenin) content of edible tropical plants, *Journal of Agric Food Chem*., vol. 49, 2001, pp. 3106-3112.

[31] G.Sadler, J.Davis, D.Dezman, Rapid extraction of lycopene and B-carotene from reconstituted tomato paste and pink grapefruit homogenates, *Journal of Food Science*, vol.55, no. 5, 1990, pp. 1460-1461.

[32] S.K.Yadav, A.Sehgal, Effect of home processing on ascorbic acid and beta-carotene content of spinach (Spinachiaoleracia) and amaranth (Amaranthus tricolor) leaves, *Plant Foods for Human Nutrition*, vol.47, 1995, pp. 125–131.

[33] S.K.Yadav, A.Sehgal, Effect of home processing on ascorbic acid and beta-carotene content of bathua (Chenopodium album) and fenugreek (Trigonellafoenumgraecum) leaves, *Plant Foods for Human Nutrition*, vol. 50, 1997, pp. 239–247.

[34] O.Fennema, Loss of vitamins in fresh and frozen foods, *Food Technology*, vol.31, no.12, 1997, pp. 32–38.

[35] Price K. R., Casuscelli F., Colquhoun I. J., Rhodes J. C., Composition and content of flavonol glycosides in broccoli florets (Brassica olearacea) and their fate during cooking, *Journal of the Science of Food and Agriculture*, vol.77, 1998, pp. 468–472.

[36] Tiong Ngee, K.Wen, Nagendra Prasad, Bao Yang, Amin Ismail, Bioactive substance contents and antioxidant capacity of raw and blanched vegetables, *Innovative Food Science & Emerging Technologies*, vol.11, no. 3, 2010, pp. 464–469.

[37] F.Shahidi, M.Naczk, Food phenolics — sources, chemistry, effects, applications. Chapter 1: Food phenolics: an overview, and Chapter 8: Antioxidant properties of food phenolic, *Techonomic*, PA, USA, 1995.

[38] P.C.H.Hollman, M.G.L.Hertog, M.B.Katan, Analysis and health effects of flavonoids, *Food Chem.*, vol. 57, 1996, pp. 43–46.

[39] R.J.Nijveld, E. van Nood, D.E.van Hoorn, P.G.Boelens, K.van Norren, P.van Leeuwen, Flavonoids: a review of probable mechanisms of action and potential applications, *American Journal of Clinical Nutrition*, vol. 74, 2001, pp. 418–425.

[40] M.Mayeaux, Z.Xu, J.M.King, W.Prinyawiwatkul, Effects of cooking conditions on the lycopene content in tomatoes, *Journal of Food Sciences*, vol. 71, no.8, 2006, pp.461–464.

[41] K.A.Thompson, M.R.Marshall, C.A.Sims, C.I.Wei, S.A.Sargent, J.W.Scott, Cultivar, maturity and heat treatment on lycopene content in tomatoes, *Journal of Food Sciences*, vol. 65, 2000, pp. 791–795.