

ISSN 2029-249X (spausdintas)
ISSN 2029-4824 (internetinis)

VYTAUTO DIDŽIOJO UNIVERSITETAS
KAUNO TECHNOLOGIJOS UNIVERSITETAS
VILNIAUS UNIVERSITETO KAUNO HUMANITARINIS FAKULTETAS

INFORMACINĖS TECHNOLOGIJOS

XXI TARPUNIVERSITETINĖS TARPTAUTINĖS MAGISTRANTŲ IR DOKTORANTŲ
KONFERENCIJOS „INFORMACINĖ VISUOMENĖ IR
UNIVERSITETINĖS STUDIJOS” (IVUS 2016) MEDŽIAGA

2016 M. BALANDŽIO 28 D., KAUNAS, LIETUVA

VYTAUTO DIDŽIOJO UNIVERSITETAS
2016

Redakcinė kolegija

Tomas Krilavičius, Vytauto Didžiojo universitetas, Lietuva
Robertas Damaševičius, Kauno technologijos universitetas, Lietuva
Arūnas Lipnickas, Kauno technologijos universitetas, Lietuva
Audrius Lopata, Vilniaus universiteto, Kauno humanitarinis fakultetas, Lietuva
Aušra Mackutė-Varoneckienė, Vytauto Didžiojo universitetas, Lietuva
Ka Lok Man, Xi'an Jiaotong-Liverpool universitetas, Kinija
Vaidas Morkevičius, Kauno technologijos universitetas, Lietuva
Sigitė Pečiulytė, Vytauto Didžiojo universitetas, Lietuva
Rasa Ruseckaitė, Monash universitetas, Australija
Inguna Skadina, Latvijos universitetas, Latvija
Lucio Tommaso De Paolis, Salento universitetas, Italija
Minija Tamošiūnaitė, Giotingeno universitetas, Vokietija ir Vytauto Didžiojo universitetas, Lietuva
Audrius Varoneckas, Vytauto Didžiojo universitetas, Lietuva
Ivan Zapreev, ASML, Olandija
Indrė Žliobaitė, Aalto universitetas, Suomija

Konferencijos organizacinis komitetas

Pirmininkas: Tomas Krilavičius, Vytauto Didžiojo universitetas, Lietuva
Andrius Davidsonas, Vytauto Didžiojo universitetas, Lietuva
Edgaras Ščiglinskas, Vytauto Didžiojo universitetas, Lietuva
Paulius Stankevičius, Vytauto Didžiojo universitetas, Lietuva
Robertas Matusa, Vytauto Didžiojo universitetas, Lietuva
Rūta Užupytytė, Vytauto Didžiojo universitetas, Lietuva
Vytautas Mickevičius, Vytauto Didžiojo universitetas, Lietuva

Leidinio redaktorius

Audrius Varoneckas, Vytauto Didžiojo universitetas, Lietuva

ISSN 2029-249X (Print)
ISSN 2029-4824 (Online)

VYTAUTAS MAGNUS UNIVERSITY
KAUNAS UNIVERSITY OF TECHNOLOGY
VILNIUS UNIVERSITY KAUNAS FACULTY OF HUMANITIES

INFORMATION TECHNOLOGY

PROCEEDINGS OF THE XXI INTERNATIONAL MASTER AND
PHD STUDENTS CONFERENCE “INFORMATION SOCIETY
AND UNIVERSITY STUDIES” (IVUS 2016)

28 APRIL 2016, KAUNAS, LITHUANIA

VYTAUTAS MAGNUS UNIVERSITY
2016

Editorial board

Tomas Krilavičius, Vytautas Magnus University, Lithuania
Robertas Damaševičius, Kaunas University of Technology, Lithuania
Arūnas Lipnickas, Kaunas University of Technology, Lithuania
Audrius Lopata, Vilnius University, Kaunas faculty of Humanities, Lithuania
Aušra Mackutė-Varoneckienė, Vytautas Magnus University, Lithuania
Ka Lok Man, Xi'an Jiaotong-Liverpool University, China
Vaidas Morkevičius, Kaunas University of Technology, Lithuania
Sigitas Pečiulytė, Vytautas Magnus University, Lithuania
Rasa Ruseckaitė, Monash University, Australia
Inguna Skadina, University of Latvia, Latvia
Lucio Tommaso De Paolis, University of Salento, Italy
Minija Tamošiūnaitė, University of Göttingen, Germany and Vytautas Magnus University, Lithuania
Audrius Varoneckas, Vytautas Magnus University, Lithuania
Ivan Zapreev, ASML, The Netherlands
Indrė Žliobaitė, Aalto university, Finland

Steering committee

Chair: Tomas Krilavičius, Vytautas Magnus University, Lithuania
Andrius Davidsonas, Vytautas Magnus University, Lithuania
Edgaras Ščiglinkas, Vytautas Magnus University, Lithuania
Paulius Stankevičius, Vytautas Magnus University, Lithuania
Robertas Matusa, Vytautas Magnus University, Lithuania
Rūta Užupyte, Vytautas Magnus University, Lithuania
Vytautas Mickevičius, Vytautas Magnus University, Lithuania

Proceedings editor

Audrius Varoneckas, Vytautas Magnus University, Lithuania

TURINYS / CONTENT

L. Mikelsonis, E. Karčiauskas APPROACH TO EVALUATION OF CORRELATION CODING TIME WITH SOFTWARE METRICS	6
I. Lagzdinytė-Budnikė, J. Sobolevas CENTRALIZED NETWORK ACCESS CONTROL SYSTEM BASED ON THE SPATIO-TEMPORAL-RBAC MODEL AND RADIUS PROTOCOL.....	11
M. Liutvinavičius, J. Zubova, V. Sakalauskas FINANSŲ RINKŲ PROGNOZAVIMAS REMIANTIS INVESTUOTOJŲ NUOTAIKŲ INDIKATORIUMI	17
E. Burokas, D. Barisas COMPARISON OF POPULAR PHP FRAMEWORKS	23
R. Ramonas, V. Moskaliova FINANSINIŲ PIRAMIDŽIŲ MODELIAVIMAS.....	28
R. Mateika OSCILLATOR GENERATED SIGNAL STRENGTH EVALUATION METHOD	32
G. Ciganaitė, T. Krilavičius, A. Mackutė-Varoneckienė LITHUANIAN TEXT DOCUMENT PROCESSING AND CLUSTERING	36
D. Šopa DVIMAČIŲ SUPJAUSTYMO UŽDAVINIŲ SPRENDIMAS NAUDΟJANT GRAFINĮ PROCESORIŪ	40
Š. Slanys, Š. Packevičius RESEARCH AND IMPLEMENTATION OF TASK AND PROJECT MANAGEMENT SYSTEM BASED ON EXT JS FRAMEWORK.....	45
A. Čiuksys, R. Butkienė ASAS – LIETUVIŲ KALBOS ANAFORŲ SPRENDIMO ANALIZĖS IR ĮVERTINIMO PROTOTIPAS	49
A. Indriulionis, J. Bielskus, P. Vaitkus NEW RADIAL BASIS FUNCTION METHOD MODELLING BOREHOLE HEAT TRANSFER: THE THEORETICAL SURVEY	55
L. Valiukonytė, L. Bisikirskienė ORGANIZACIJOS MODELIŲ DUOMENŲ PAKARTOTINIS PANAUDOJIMAS VEIKLOS PROCESU MODELIAVIME	61
I. Sedunovas MAŠINŲ, TEIKIANČIŲ INFORMACIJĄ MTCONNECT PROTOKOLU, DUOMENŲ SURINKIMAS IR ANALIZĖ	66
A. Nauckūnas, E. Sakalauskas MOBILE IDENTIFICATION TO VPN SYSTEM USING CRYPTO PROCESSOR AND NFC	72

Approach to evaluation of correlation coding time with software metrics

Laimonas Mikelsonis

Kaunas University of Technology
Faculty of software engineering
Kaunas, Lithuania
Laimonas.mikelsonis@ktu.edu

Eimutis Karčiauskas

Kaunas University of Technology
Faculty of software engineering
Kaunas, Lithuania
Eimutis.karciauskas@ktu.edu

Abstract - This paper shows how time spent on software developing correlates with metrics of software engineering. This leads to analysis of metrics of software engineering, what these metrics can describe and how it can describe quality of software. This is because each stage of software developing requires time. This will lead to correlation between concrete code metric and time spent on it. This paper describes how time calculation could be improved and collected data could be used for estimating projects in the future. This leads to estimating development time based on predicted value of metrics.

Keywords - Time, Software Quality, Software Metrics, Estimates

I. INTRODUCTION

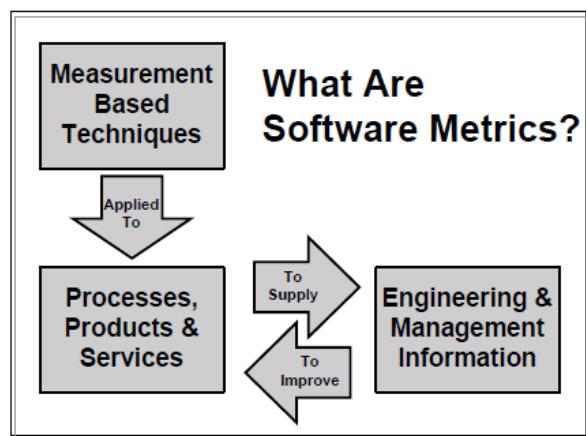
The idea of this article was to check how time is related with software metrics that are used by many companies and developers. This research helps to check how to evaluate your work, allows checking your developing speed and etc. Main goal of this experiment was to find connection between time and other basic software metrics and to have a method to collect data for future development and estimation. For this purpose the analysis of software metrics was made. Explaining of what it is, how to work with it and what advantages and disadvantages it gives for developer/customer/company. The experimental part is a description of improvement for existing time estimate methods. The result allows checking your developing speed, checking how the complexity of the software changes in the period of time. With collected data and estimated project size estimation of project time can be made. This allows estimating project times more accurate.

II. DESCRIPTION OF SOFTWARE METRICS

Code metrics is one of the part of software engineering process that describes software quality and complex of software. As Linda Westfall explains “Software metrics are an integral part of the state-of the-practice in software engineering. More and more customers are specifying software and/or quality metrics reporting as part of their contractual requirements. Industry standards like ISO 9000

and industry models like the Software Engineering Institute’s (SEI) Capability Maturity Model Integrated (CMMI®) include measurement. Companies are using metrics to better understand, track, control and predict software projects, processes and products.”[2]. This statement describes how we should understand software metrics, what it is and how it is implemented in such industry standards like ISO 9000.

Nowadays customers can include software metrics in their requirements because it may help to improve quality and maintainability for software. But there is one of the challenges for software metrics because there are only few standardized mapping systems. This means that such software metrics like lines of codes (LOC) does not have standard measurement methods [2]. For instance, when measuring the lines of code we should decide how we will count them: “Do we count physical or logical lines of code? Do we count comments or data definition statements? Do we expand macros before counting and do we count the lines in those macros more than once?”[2].



This is an abstract statement as we cannot have only one internationally accepted method for counting lines of code for software.

Figure 1. What are software metrics? [2]

As this article talks about time and other software metrics we must know how we can measure time for projects, for example, engineering hours. This is also very hard to

calculate metric because there can be a question: “[...] do we include the effort of testers, managers, secretaries, and other support personnel” [2]. Although there are two metrics that has concrete and internationally accepted method to measure – it is McCabe’s Cyclomatic Complexity and Function Point Counting Standard. It is described by International Function Point User Group (IFPUG).

When focusing on time measurement we should go through some steps and make a model how we can calculate the time and how we can connect it with other software metrics. First, we must define customer. Customer can be actual customer of software or one of the software developing team. For example: testers, managers or developers. We want to calculate time spent on developing and its correlation to other software metrics. This would be good for developers and managers. Managers could give report for customer about time spent on the software developing and progress of it. When customer is identified we have to define our goals - where these metrics results will be used and how. “Basili and Rombach define a Goal/Question/Metric paradigm that provides an excellent mechanism for defining a goal-based measurement program. Figure 2 illustrates the Goal/Question/Metric paradigm.”[2]

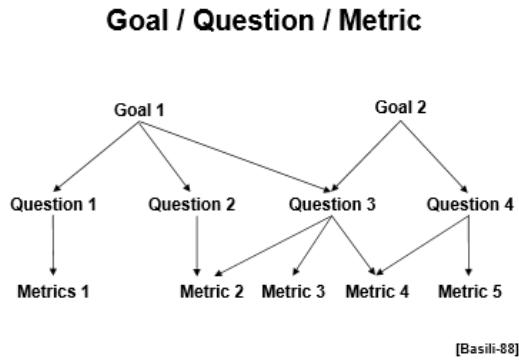


Figure 2. Goal/Question/Metric [2]

In this case our goal is to measure each software development stage time and how the code changes in measured time. This article explains not only time as a software metrics but also how it interacts with other metrics. We should choose what metrics we would like to correlate with. Better and easier way is to calculate: Lines of Code, McCabe’s Cyclomatic Complexity and Function Point. These three metrics will be correlated with time spent while developing software.

Software metrics, that needs to be measured and correlated, is selected, but all metrics must have standard definitions [2]. Units how these metrics will be calculated must be selected. For this step it is good to use standard units. Lines of codes are measured by lines per file (class). McCabe’s Cyclomatic complexity has his own algorithm of how it must be measured.

Functional Points, in other words, is just functions calculated per file. Time must be measured in seconds.

There is an example of how time is expressed in other units, also as a metric of software. Alain Abran described about Halstead’s metrics. His article explains that “Halstead’s Metrics are commonly known collectively as ‘software science’ [1]. “Researchers have used them:

- To evaluate student programs and query languages
- To measure software written for a real
- time switching system,
- To measure functional programs,
- To incorporate software measurements into a compiler, and more
- recently
- To measure open source software” [1].

These metrics measurement principles and metrics are included in number of current commercial tools that are used to count lines of code. According to Alain Abran, Halstead found a function how to express time as software metric. Halstead defined that “The required programming time (T) for a program P of effort E is defined as:

$$T = \frac{E}{S} = \frac{n_1 N_2 N \log_2 n}{2n_2 S}$$

where S is Stroud number...[1] “and the E is s defined as a measurement of the mental activity required to reduce a preconceived algorithm to a program P”[1]. N is sum of N_1 and N_2 . And n is equal to $n_1 + n_2$. Halstead defined that all base metrics can be measured using these.”

- n_1 : Number of distinct operators.
- n_2 : Number of distinct operands.
- N_1 : Total number of occurrences of operators.
- N_2 : Total number of occurrences of operands”[1]

This allows expressing required time as a function. The Stroud number used in the function can be described as “The Stroud number S, indicating the number of elementary decisions per second, satisfies this definition. Stroud’s number has values between 5 and 25 elementary discriminations / second, 25 being the upper limit of images people can discriminate within a second.”[3]. Halstead make a description of how the complexity of the software or program can be measured by having “measurable properties of the program. One of those properties is the language’s vocabulary (number of distinct characters and signs, or operators and operands).” [3]. This gives definition about Stroud number. According to Stroud, in software science Stroud number is set to 18.

III. AGILE AND STATISTICAL METHODOLOGY

Agile methodology has its own rules for measuring and estimating time for software development. Differently from waterfall method, Agile suggests doing all tasks in iterations. Iterations are a time frame in which some stories (in Agile stories means tasks) must be done. Length of iteration can vary on team and company but usually it is two weeks. In that time team/developer should finish the list of tasks that is

assigned for this iteration. Stories must have a point value which represents difficulty of the story and time, which will be spent while developing software. Usually, there are three point scales: 1, 2 and 3 points. Each team/developer can choose its own system to set these points. For instance, some of the team estimates time for story using Fibonacci values: "so a very small story will be one point, a slightly larger one, two points, a slight larger one than that three points, then five points, eight points, 13 points, etc." [5]. This helps to know how many story's points did the team finish in the iteration. At the beginning of Agile all teams will fail. This is because they have to do some of iterations and get a number of point values of the story that can be done in iteration. For instance, if a team at the beginning thinks that they can do 60 points in one iteration, they will fail and do for example 40. Then next time the team can predict point for iteration carefully and more accurate. After some time, team can become good and accurate Agile team. But this means that we know only point for the story and how many points we can do in one iteration. There is no meaning for estimates. There comes another way to include time into Agile methodology and to estimate time for the software development. The system can be called "ideal hours" [5]. This can be described as "A team that estimates using ideal hours thinks, 'If we had no meetings, no appointments, and no distractions at all, how many hours would it take to finish this story?'" [5]. Then team will track how many hours they have spent to finish the story and will calculate their velocity for the next iterations. However, "ideal hours" method is very optimistic but it never works. Typically, real hours will be half of the ideal hours. Even best of Agile teams cannot do better than ratio between real hours and "ideal hours" more than 3:4 [5]. In this case, the time estimate in Agile methodology can be done in many ways and it depends on the team/developer which way they are going to use. But when the team has velocity they can estimate times for next stories/project very accurate. But this is quite new and fresh Agile methodology. There is another way to estimate times not using Agile.

The statistical method explains more about why it is not possible to estimate time accurately:

- The productivity and experience level of the engineer, particularly if multiple people are involved
- PTO, Late arrivals, early departures, sickness, etc.
- Unforeseen defect and customer requests, troubleshooting, challenges, system/environment issues, software/library issues, learning and ramp-up, design/architecture, required research, etc.
- The fact that software engineers notoriously underestimate
- Unforeseen issues with maintainability, architectural flaws/imperfections, scalability, performance, testability, etc.
- Time associated with spikes, R&D, design, architecture, mockup, prototype, POC, etc.
- Administrative work and non-engineering related requests" [4]

This reveals the problem that causes inaccurate time estimation for software development process. Some people think that "One may think that the only way to get a better estimate of time and delivery date is to try to subtask and think of everything **up front**, which involves more requirements gathering and documentation, mockups, prototypes, UML diagrams (sequence, use case, etc.), and so on." [4]. This method is opposite to Agile. This method is also wrong because of requirements. They can change before development start, for example, on testing level, on beta test, etc. This will lead to an inaccurate estimated time. Time estimate is very important in software engineering and it may cost a lot. There is theory that developers spent 5% of their time working with task related with estimating the time. It can cost for developer/company/team a huge amount of money. How can we get rid of problems with inaccurate estimates but not use Agile? There is another approach. As Alex Castrounis described "that I use for characterizing the relative size of development tasks is a variation of the tee-shirt sizing method. Each task is given a relative size corresponding to five tee-shirt sizes, along with a very rough lead time estimate (for a single developer) as shown.

- XS: Half day or less
- S: Half day to one day
- M: Two to three days
- L: One week
- XL: One to two weeks

"[4]. Tasks longer than two weeks must be split into subtasks. This will lead to more accurate estimation of time and also more data collected to statistical methods, which will help to improve accuracy of estimate by using history of tasks with the same size.

Similar research to this article has been made by Wolfgang Holz, Rahul Premraj, Thomas Zimmermann, and Andreas Zeller. They made an assumption that they can predict software metrics by includes in the code [6]. They used 89 Eclipse plugins and made a *predictor* – a prediction software that predicts metrics. For example, this program can predict source lines of code by includes in the code. This prediction is made by using machine learning algorithm. They give a lot of data to analyze and predictor is *learning*. After that they ask *him* to predict software metrics for next project by number of includes. As mentioned in this article, this can be improvement for existing methods. In *predictor* method they can predict metrics and this described method can show correlation between metrics. Then *predictor* methods can also predict time for development. For example, if they predict that software size will be 10000 lines of code and described methods show that developer can write 10000 lines of code in 2 weeks then estimate for the project can be made.

There is also another model that can be used for predictions and estimations – COCOMO [7]. This method is also theoretical and can estimate efforts, duration, size and etc. of the software. This leads to the same conclusion as mentioned above. We can make prediction of how many lines will be used in the code, according to: inputs for the software, predicted number of lines of code, requirements type, methods

and etc. This can predict duration and efforts of software, but it is in theoretical approach. COCOMO method can be connected with the method described in this article, using data collected through time, and using statistical methods - more accurate estimations can be done. This is because real practical data is more accurate than assumption.

IV. DATA FOR THE METHOD

Method that is going to be described in this section is a suggestion for existing methods, as an improvement for correct time estimates, and correlation between other software metrics. Method is based on both previously described methodologies – Agile and statistical. The idea of this analysis is to collect data from previous task and to make a calculation to try to estimate next project/task/story time.

It is a part of statistical method because it is using time values for stories/projects or files and tells how much time you have spent on other projects, task or file (project class file or etc.). This will allow collecting data and using statistical methods and some calculations predict time for next project.

Agile methods come here too. As described above, Agile is based on giving stories (tasks) point and knowing how many points developer/team can complete in one of iteration. So from some perspectives this is also based on statistics. If in Agile, developers give stories' points; in this method other software metrics can be assigned to the story. For example, task is to write function to get data from database. Method can calculate time, cyclomatic complexity and functional points for the story and save it into database. Next time when similar story comes, collected data can be analyzed and estimation for a new story can be made. In that time, accuracy of estimation can be improved because every time when the story is completed new data comes to the database.

For this method there is some data that must be collected to make method work. It is described in the table below (Table 1).

Data type\Column name	ID	ProjectName	ClassName	TimeSpent	CyclomaticComplexity	FunctionalPoints	DateTime	TaskNumber	LineOfCode
	integer	string	string	int	int	int	DateTime	int	int

Table 1. Data that must be collected for methods.

As we can see by the data description in the table there is a lot of metrics that must be selected.

ID – it is unique identifier for record in the database.

ProjectName – every story/task depends to project.

ClassName – concrete file in which we are going to write code.

TimeSpent – calculated time for how long developer was working in this particular class.

CyclomaticComplexity – this metrics must be calculated using some methods and function or some third party software. But this should be done in IDE while working.

FunctionalPoints – this metrics, also as cyclomatic complexity must be calculated using methods or software.

DateTime – date and time when the record was inserted.

TaskNumber – if task/story that we working on has number it can be saved also here to map data for story. This allows you to map multiple records in database to one story/task.

LineOfCode – this also must be calculated using some methods or software.

Best practice to collect data is to develop extension of software development IDE and make it automatically collect data for specific tasks. This will lead to correct data and better accuracy in analysis stage.

When data is collected there can be time calculated that was spend on that task that we have in the database. This allows to sum all time. Then, exact time which was spent on the story can be entered in project management system. This will give some statistical data for future development. If the task for example is rejected or must be done in another way there can be an estimation made based on previously collected data.

V. THEORETICAL EXPERIMENT

Let's assume that we have several lines in database. For example (Table 2):

ID	PrN	Class	Time	CC	FP	DT	TN	LOC
0	Test	Class1	15	2	2	02/02/16	1	54
1	Test	Class2	25	1	1	01/02/16	2	36
2	Test	Class1	5	3	1	03/02/16	1	40
3	Test1	Test1	26	2	1	04/02/16	11	69
4	Test1	Class5	105	5	4	04/04/16	7	105

Table 2. Example data to explain method

According to this data, we can check that there are two projects that were worked on and few different classes. There can be several conclusions done by this data.

First, data of several tasks are in database. This means that we can measure software metrics and time for these tasks. We can see that for task number 1 there are two records. For the first time developer added some code which is 54 lines length, cyclomatic complexity is 2 and number of functional points is 2. Then developer did some changes in the code. After those lines of codes decreased by 14 lines, cyclomatic complexity has increased by one and functional points have decreased by one. There is really important to check datetime value in database, because latest data is the real one. Other one is change history but it helps to see time and other software

metrics changed through development. Below there is example how data can be used for statistics (Figure 3).

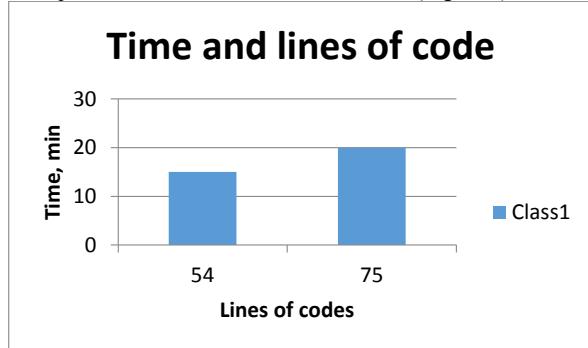


Figure 3. Time and lines of code correlation

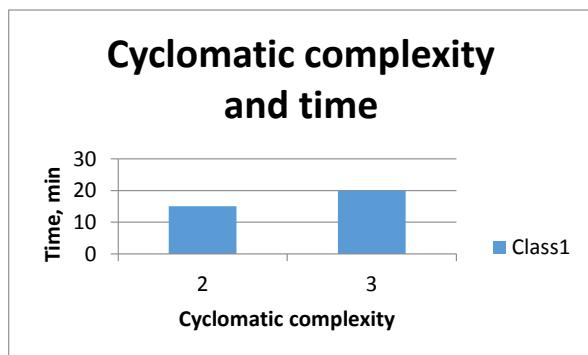


Figure 4. Cyclomatic complexity and time correlation

So for class Class1 whole time spent is 20 minutes and the total line of code in that file is 75. We can do prediction that this will also work with other software metrics and data.

VI. RESULTS THEORETICAL OF EXPERIMENT

As we can see in the diagrams there is quite easy to see how software metrics is changing by the time spent for the project/task/class. This means that when we have big amount of data collected we can calculate time. Next time when task is given, by the requirement we can predict such things as cyclomatic complexity and by that estimate time for task and etc.

VII. CONCLUSIONS

- Analysis part shows that there are a lot of methods to estimate time for tasks/project but the accuracy of these methods are different.
- Each project management methodology has its own methods to estimate time for the project and check correlation between other software metrics.
- Improvement for existing methods can help to estimate time for project/task more accuracy and this will help developers/teams/companies to save a lot of time and money.
- Using already existing methods to estimate software size and connection with described method statistical data can lead to very accurate estimation for the project time.
- Software must be developed to prove this theoretical example.

VIII. REFERENCES

- [1] Alain Abran. "Software Metrics & Software Metrology". 2010. [Reviewed 2016 02 22] Available: <<http://prof.etsmtl.ca/abran/Accueil/ChaptersBook/Abran%20-%20Chapter%20007.pdf>>
- [2] Linda Westfall "12 Steps to Useful Software Metrics". [Reviewed 2016 02 19] Available: <http://www.westfallteam.com/Papers/12_steps_paper.pdf>
- [3] Cees Jan Koomen, "The Stroud number in engineering" [Reviewed 2016-02-22] Available: <http://www.eetimes.com/author.asp?section_id=36&doc_id=1265859>
- [4] Alex Castrounis, "Why Software Development Time Estimation Doesn't Work and Alternative Approaches" 2015-08-15. Available: [Reviewed 2016-02-22] Available <<http://www.innoarchitech.com/why-software-development-time-estimation-does-not-work-alternative-approaches/>>
- [5] Chris McMahon, "Estimation approaches in Agile development". 2011 May. [Reviewed 2016-02-23]. Available: <<http://searchsoftwarequality.techtarget.com/tip/Estimation-approaches-in-Agile-development>>
- [6] Wolfgang Holz, Rahul Premraj, Thomas Zimmermann, Andreas Zeller. "Predicting Software Metrics at Design Time". [Reviewed 2016-04-12]. Available: <<http://thomas-zimmermann.com/publications/files/holz-profes-2008.pdf>>
- [7] "Overview of COCOMO". [Reviewed 2016-04-12]. Available: <<http://www.softstarsystems.com/overview.htm>>

Centralized network access control system based on the Spatio-Temporal-RBAC model and Radius protocol

Ingrida Lagzdinyte-Budnike
Kaunas University of Technology
Kaunas, Lithuania
Ingrida.lagzdinyte@ktu.edu

Jevgenijus Sobolevas
Kaunas University of Technology
Kaunas, Lithuania
Eugenijus07@gmail.com

Abstract — Paper describes designed and developed centralized network access control system based on time (temporal) and location (spatial) information – an extended RBAC model. RBAC model extensions also are related to system functioning in wireless network specifics. The system allows to distribute access rights to the wireless network users based on the location and time of network access, authorization result for network user depends on actual location and current date/time. The system supports Radius protocol, which provides compatibility with various network devices. System has modular architecture that makes it easier to extend or replace existing functionality. System operations are described using context and sequence diagrams.

Keywords — Radius; STRBAC; Centralized; Wi-Fi; Wireless network; access control; RBAC; location based authorization; time based authorization;

I. INTRODUCTION

Nowadays wireless networks are widespread and their usage is steadily increasing [1, 2]. Wireless networks are used privately, at home, at public areas (shopping malls, restaurants, outdoors), in the educational and administrative authorities and etc. The spread of wireless networks increases a prevalence of hacking or unauthorized access risks, thus complicating an access control. Access to wireless networks is not always properly protected, secured and managed [3].

Widespread of the wireless networks and wide usage of the mobile devices creates the need for such applications and services that can use time and location information to improve their existing functionality. For example, in the university during exams there would be convenient to have the possibility to deny access to the wireless network for the group of students who are passing exam only for the classrooms where exam is hold. At same time the access restrictions would not be applied for the lecturers in the same classrooms. The access control can be accomplished even if students and teachers are connected to wireless network via same access point. The software can be useful and for business entities, i.e. hotels providing wireless network access for certain customers in the hotel rooms. Similar type of functionality can be useful or even important for the restaurants, banks, various public events and areas with the

requirement to allocate access to wireless network by location, limiting area of permitted access for relevant users, including allowed access time information. User location information and connection time can be used to improve protection of network access control.

To support above mentioned functionality it would be required more complex access control systems. There is growing need to modify existing access control models to support advanced features.

One of the most commonly used practices for the standard access control implementation is based on the role-based access control (RBAC) model [4, 5]. Using this model, a role is assigned to each user, and then permissions are set to each role. In such case different users can have different access rights.

There are several extensions of the RBAC model that allow including of the additional information about the users in the process of the authorization [6-9].

There are various studies and researches that focus on possibilities of RBAC model extensions [10, 11]. In these studies, authors make focus on the RBAC extension on the conceptual level, but in the most cases they are not providing at all or providing very little information about how the following models should be adopted to the different environments or integrated with other systems, i.e. the model adaptation to use with the wireless network.

Further in this paper similar other researches are discussed in chapter II. The architecture and basic principles of our implementation of the centralized access control system based on STRBAC model are described in chapter III. Testing process of the developed software and further work required for the system improvement, including parameters optimization tasks are specified in chapter IV.

II. RELATED WORKS

Similar researches have been executed by Yue Zhang and his team. They offer RBAC model extension with time information (called “Generalized Temporal Role Based Access Control - GTRBAC”). In the research work of

Abdunabi and authors RBAC model was extended with space and time information, then it was used in smart phone application “iMedik” [12, 13]. Further these two case studies will be discussed.

A. Deployment of the GTRBAC model

Yue Zhang and James B.D. Joshi studied security issues of the various resources and were trying to extend regular RBAC model with temporal information. They proposed additional RBAC model extension which supported rules and authorization based on the temporary information. Studies have shown that such functionality and features greatly facilitates system administration. However, GTRBAC model quite rarely is used in real environments, because of complexity in the implementation stage. Authors have mentioned, that GTRBAC model has a lot of constraints, which should be taken into account and makes implementation challenging and complicated, especially when it is applied to the real world environments.

In GTRBAC model we can distinguish the following variable constraints:

- *Time* – the ability to assign permissions or restrict user actions by the time. For example, the employee’s working hours are from 9 a.m. to 5 p.m. At these time employers can use system resources, after work hours the system will deny access.
- *Periodicity* – the possibility for the periodical authorizations or restrictions. For example, an employee uses the system only on Mondays. On other days access might be restricted.
- *Quantity or quantitative restriction* – allows to set the number of how many times the certain resource or service can be used.

Above mentioned constraints define restrictions, which are used for the allocation of allowance. Using GTRBAC model co-authors have developed a prototype system, which allows to distribute permissions and restrictions depending on the temporal information. However, user location information usage for RBAC model was not discussed.

B. GSTRBAC model usage in mobile application “iMedik”

Ramadan Abdel Nabi and co-authors were solving how to improve private information protection of clinic patients. Doctors in the clinic used mobile and handheld computers with “iMedik” application which allowed to observe personal information like disease description, home address and etc. However, there was a problem – how to protect patients’ personal information in such a way, that doctors could reach personal data of the patients only in the clinic place and only during working hours. Doctor can lose a mobile device with all patients’ personal data (online data can be saved), and any other ineligible person who will find lost mobile device would be able to access patients’ personal information and use it by his own needs. Traditional role based access control models, such as RBAC does not solve such problem.

The authors’ proposed solution was GSTRBAC model that takes into account not only the user’s role, but also the location and time information. During the attempt to reach personal information of the patient, the time and location of the doctor is being checked. If the doctor or other person finds the lost mobile device and attempts to access the patient’s information outside the clinic or after working hours he would be blocked. To identify location of the user, authors have used GPS coordinates. Therefore, GPS is not the best solution for location determination *inside buildings*.

III. DEVELOPED SYSTEM DESCRIPTION

The main system goal is to improve the wireless network authorization process by allowing to authorize network users depending on their location and connection time. The user location is obtained by other external system. To achieve that it was decided to use RBAC model concepts:

- a) By extending RBAC with the location and time information;
- b) By evaluating if extended RBAC can be applied on the functions of the wireless network specifics;

A. Spatio Temporal Role Base Access Control (RBAC) model

The RBAC model extended with time and location information (STRBAC) is shown below in Fig. 1. The components of the model are:

- *User* – a person or device that will use the system.
- *Role* – right or responsibility that is assigned to the user (e.g. any work function in the organization).
- *Permission* – right or permission to perform an action with an object or resource.
- *Time information* – extends the role or permission with time information that allows configuring permissions or restrictions for the user actions depending on the time.
- *Periodicity* – allows configuring restrictions or permissions periodicity (e.g. on each Monday).
- *Location information* – allows to configure restrictions or permissions depending on the location of the user. The user’s location is obtained by another system, which informs developed system by providing to it exact room name where user is located. For example, the user A is in 302 classroom. We know that the 302 classroom is located on the 3rd floor. So it can be said that location detection supports $x;y;z$ coordinates.

Using such model each user is linked with a certain role (permissions), which is then linked with time and location information. In addition, such permissions may be periodic, e.g. user A can use resource $R1$ every Monday at 2 p.m. from location $L1$.

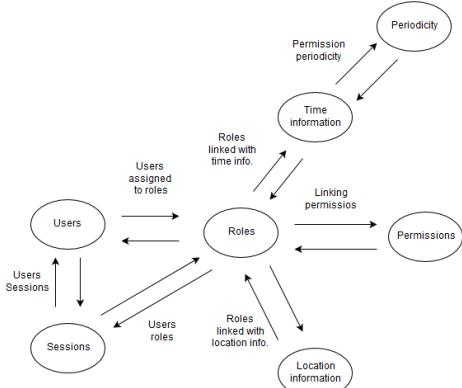


Figure 1. “STRBAC model”

This model allows dividing users into groups (roles), where each group is linked with the specific access rights. Moreover, permissions are linked with permissive access, time and location information, as well as the periodical permits. For example, employees who belong to group $G1$ will be able to use the network from the place $L1$, only on Mondays and on Fridays from 8 a.m. to 5 p.m. The access on all other days will be denied. However, the group $G2$ members will be able to use the network from $L1$ on any day.

B. Developed system functionality

System management is made through a graphical user interface, which allows:

1. *Manage network users.* Network user management includes: creation of the user, user personal information editing, deleting existing user.
2. *Manage user groups.* Management of the user groups (groups corresponds to roles) includes: assign or remove a user to or from one or more groups. Creating new or deleting existing group, changing group description. Group management includes linking of the group with permitted network access places and time profiles. Each group optionally can be linked with the notification template.
3. *Event log.* System tracks all important events such as administrator actions and system actions, including actions of the network user (authorization, accounting, disconnecting), sent notifications and other information. Event log is provided within graphic user interface with search and different filters features. On each log event it provides additional information about events, such as why the network user is prohibited to grant network access or why the user has been disconnected from the network.
4. *Update of the building, location plans or places.* Location information of the network users is requested from other external system (positioning service). In order to ensure smooth communication between developed authorization and positioning systems,

building plans in both systems should be synchronized. If the location maps (contains all buildings, floors and rooms) are updated in positioning service, then they should be synchronized with the developed authorization system. This action can be triggered from graphic control panel and processed automatically. Update of location maps in developed system consists of removing the outdated locations and downloading, updating, saving the updated ones. Such process ensures that location objects in both systems are synchronized.

5. *Manage time profiles.* Time profiles management consists of creating a new profile, editing current profile and removal of existing profile. Time profiles enable identification of permissive connection time range and weekday for the certain location. The system allows to link different time profiles with multiple locations which are also assigned with certain group. For example, the group $G1$ members are allowed to use network from room $R1$ only on Mondays from 9 a.m. To 5 p.m., and the $G2$ group members are allowed to use network from the same room all day long.

6. *Manage SMS notifications templates.* System provides notifications via SMS, custom notification template can be defined for each supported notification type. SMS notifications can be sent to the system administrator providing important information or to certain group members providing relevant information. Each location assigned to the group can have different SMS notification. For example, user $U1$ is a member of group $G1$. Notification template $T1$ is assigned to room $R1$ and $G1$. When user $U1$ will be granted to access the network and his location would be $R1$, he will get notification $T1$. It is possible to determine notification by type. The notification message can be sent when user gets right to access the network. Another notification can be sent when user is disconnected or user is connected from the specific room.

Above listed system features allows to have user groups, that could be assigned different access permissions with the exact location and time profile. Each network user can be assigned to one or more groups, location and time profile can be assigned to each user group.

Network user authorization process is fully automatic and does not require any action of system administrator. This process is explained in more detail in chapter III section D “Network user authorization process”.

C. System architecture

The system has a modular architecture where functionality can be easily extended or improved.

System was developed using Java DAO (Data Access Object) pattern [14], which allows to separate system business logic from database layer. The pattern also helps to reduce system maintenance or replacement costs.

In order to migrate existing database vendor to other, Hibernate technology was used. It ensures a smooth migration to the supported databases.

System context is shown on Fig. 2. It shows system business processes. The developed system components are marked in gray. Data between the sub systems (system management subsystem and control sub system) is transferred using RMI. Communication between target network devices, routers or access points, user authentication subsystem is organized by RADIUS protocol.

The system communicates with the SMS notifications messaging service via HTTP POST requests.

We can distinguish 3 subsystems in the system:

- 1) Network user authentication subsystem;
- 2) The control subsystem;
- 3) System control panel with graphical user interface.

Network user *authentication subsystem* is responsible for processing and analysing the RADIUS packets, network user authorization process. Management subsystem is responsible for users, user groups, locations mappings and time profiles administration.

Location positioning service is responsible for obtaining user location and providing it to our proposed authorization system. By now instead of the location system its simulator is used. Authorizaiton system communicates with location positioning system simulator via RMI. In the future proposed system can be tuned up (implemented) to work with real positioning service and communicating for example via HTTP requests or in other way. Today there are different ways to obtain indoor user location. For example, Wi-Fi based positioning system (WPS) [19], Received signal strength indication (RSSI) [20] and other ways [21, 22].

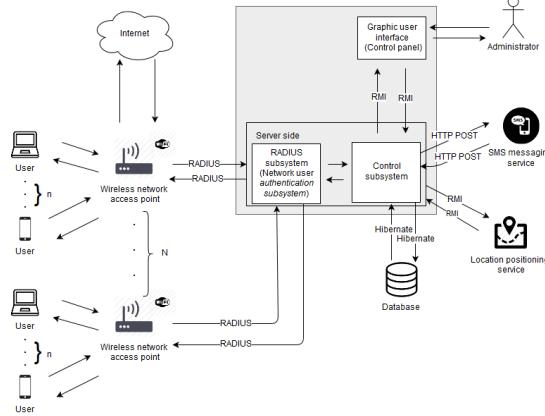


Figure 2. "System context model"

Control subsystem consists of client and server sides. Client side delivers the graphical interface, which provides functions for system management (system control panel). Server side

includes all functionality which is required to ensure system smooth work.

The control subsystem and system control panel are physically separated and may be running either on the same system node or on the different one.

Main users of the developed system are system administrators, which have a pre-configured account and can access the control panel. Wireless network users can be interpreted as non-direct system users. User credentials and pre-configured access rules for network access should present in the developed system, however users might not be aware of the developed system in place, since they are not communicating with the system directly.

D. Network user authorization process

The authorization process based on STRBAC model is wrapped by Radius protocol. This allows smooth system integration with the whole range of the different level wireless network devices (enterprise level routers or switches, as well as low level consumer equipment) which supports Radius protocol (RFC-2866, RFC-2138) [15, 16].

Network user authorization and accounting process is shown in a simplified sequence diagram (Fig. 3). System responses to the network user are:

- Access-Accept.
- Access-Reject.
- Access-Accept with further user accounting process.

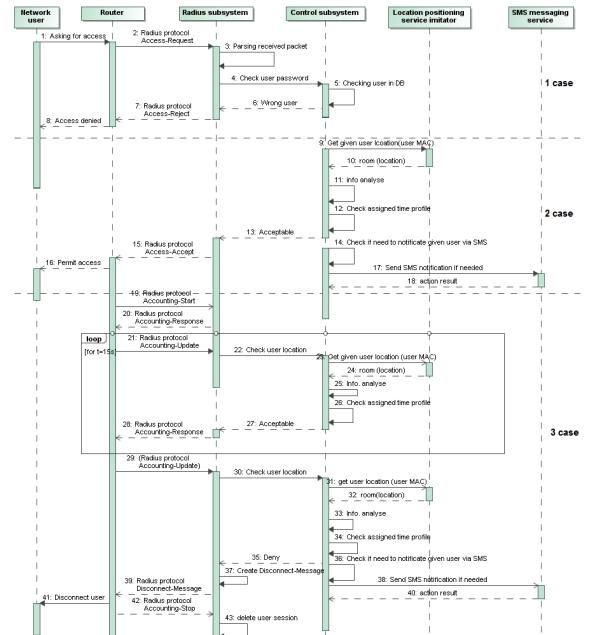


Figure 3. "Authorization process of the network users"

Initially, the network users connect to the wireless access point in usual way. Then router requests developed system authorization subsystem, to check if certain user can access the network, by sending Radius “Access-Request” packet. The authorization subsystem using control subsystem functionality checks certain user username and password in the database. If given user credentials are incorrect, system denies access, sending to router - Radius “Access-Reject” packet, access for given network user is prohibited. If the SMS notification is configured and SMS sending in the user profile is allowed (user has agreed to receive notifications via SMS messaging), he will get SMS notification explaining why his attempt to get access was rejected (Fig. 3, 1-8 steps in 1st authentication sequence diagram case).

If during authentication phase user was granted to access the network, system would provide accounting feature (in case Radius Accounting feature is supported and enabled in the router) (Fig. 3, 9-41 steps in the sequence diagram. Router sends “Accounting-Start” request and system replies with Radius “Accounting-Response”. Further periodically (time interval depends on the router configuration) router will be sending Radius “Accounting-Update” requests and on each request system will check user location and time. If it will be detected that location of the user or connection time is denied by assigned time profile to the current user group, access to the network will be prohibited. System will send back Radius “Disconnect-Message”. This packet should trigger disconnect of the appropriate user from the wireless network and inform system about accounting end (Radius “Accounting-Stop”). According to the configured notifications settings, user can be notified via SMS about any event. For example, after successful connection he can get an advertisement. If the forced disconnect occurred (by the system) from the network, he can get extra information why this has happened, i.e. room which he entered is prohibited for his group.

Accounting feature is also important if existing Wi-Fi network infrastructure is using Wi-Fi controller and fast roaming feature. In this case accounting will help to track clients migration and disconnect them from network if required. For example clients do not do reauthentication when fast roaming. Once authorized, client can move to different building if it is covered by same network. But accounting feature will check (periodically e.g. every 2m) client role (group) and assigned permissions to this role (group) if it will be noticed that this client using network at his group restricted place or time, he will be disconnected.

IV. DEVELOPED SYSTEM TESTS AND FUTURE WORKS

System testing was performed by using NTradPing [17] tool, which is designed to make and send a single Radius requests.

Unit testing automatization was made by jUnit [18] library. During unit testing individual system units and functions were tested.

Network user authorization process was tested using combined functional testing with NTradPing tool and location system simulator. For example, we can describe shortened test scenario as further: during accounting session, in the location system simulator user location was manually changing to restricted area. Simultaneously system was checking if this user was disconnected from the network. Time based restrictions were checked similarly also. Functional testing has been applied in order to test the graphical user interface (control panel).

During integration tests, several router manufacturers (*D-Link* and *Mikrotik*) were tested. All errors and inaccuracies founded during testing were successfully corrected.

In future it is necessary to create the tool that will be able to simulate the required amount of network users. Such tool will help to perform a stress test for testing average service time of network users when a lot of wireless network users are connected.

In order to fully prepare system for the effective functioning it is necessary to make following additional experimental studies:

- To detect optimal Accounting and Authorization service configurations, by analyzing servicing times and durations (average and maximal) of network users and finding correlation of those times and number of user requests (system load). Such study should show how long it may take to authenticate a single user, depending on a number of simultaneously connecting and connected to the network users.
- To identify minimal required hardware resources that are required to run system.
- Additional experimental studies might be required to give recommended time interval value for the Radius accounting requests in router configuration.

V. CONCLUSIONS

Centralized network access control system that allows authorizing network users relying on their location and weekday including time range of network access was developed and described. System provides graphical user interface that makes easier to manage users, user groups, and their access rights. System fully supports Radius protocol.

RBAC model extensions related with location and time attributes were offered and described. Model was successfully integrated and adopted to the developed system. System functioning in wireless network specifics was estimated by processing only spatial area-type attributes.

The developed system can be used for institutions, various other public places or organizations with the requirement to separate access to wireless network by location, week days, time information or to provide centralized expanded access control. System is integrated with notifications via SMS service, which not only informs network users or system

administrators about various system events, but also opens unlimited marketing possibilities.

Functionality of the system was tested using the location system simulator.

The future experimental studies should include tuning up system parameters considering different system users' location and time, including different number of system users.

REFERENCES

- [1] Office for Nation Statistics, "Internet Access - Households and Individuals", 2013. Internet access: http://www.ons.gov.uk/ons/dcp17178_322713.pdf
- [2] The Statistics Portal, "Global number of public hotspots", 2009 - 2015. Internet access: <http://www.statista.com/statistics/218596/global-number-of-public-hotspots-since-2009/>
- [3] Bogdan Botezatu, HOTforSecurity, "25 Percent of Wireless Networks are Highly Vulnerable to Hacking Attacks, Wi-Fi Security Survey Reveals", 2009 - 2015. Internet access: <http://www.hotforsecurity.com/blog/25-percent-of-wireless-networks-are-highly-vulnerable-to-hacking-attacks-wi-fi-security-survey-reveals-1174.html>
- [4] R. Sandhu, E. Coyne, H. Feinstein, and C. Youman, "Role-based access control models," IEEE Comput., vol. 29, no. 2, pp. 38-47, Feb. 1996.
- [5] Alan C. O'Connor, Ross J.Loomis, "2010 Economic Analysis of Role-Based Access Control", National Institute of Standards and Technology, December 2010. Internet access: http://csrc.nist.gov/groups/SNS/rbac/documents/20101219_RBAC2_Final_Report.pdf
- [6] Xiutao Cui, Yuliang Chen, Junzhong Gu, „Ex-RBAC : An Extended Role Based Access Control Model for Location-aware Mobile Collaboration System“, Second International Conference on Internet Monitoring and Protection (ICIMP 2007)
- [7] Yu Wanjun, Wang Yong, "Research on Security Status Recovery in Temporal Role-Based Access Control System", 2009 International Conference on Information Management, Innovation Management and Industrial Engineering
- [8] Xun Li, Sang Bong Yoo, "Extended Role-Based Security System using Context Information", 2008 Second International Conference on Future Generation Communication and Networking
- [9] Hsing-Chung Chen, Yung-Fa Huang, Syuan-Zong Lin, "Generalized Associated Temporal and Spatial Role-Based Access Control Model for Wireless Heterogeneous Networks", 2011 Fifth International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing
- [10] Elisa bertino, Barbara Catania, Maria Luisa Damiani, "GEO-RBAC: A Spatially Aware RBAC", SACMAT'05, June 1-3, 2005, Stockholm, Sweden.
- [11] Kai Ouyang, James B.D. Joshi, "CT-RBAC: A Temporal RBAC Model with Conditional Periodic Time", 2007, IEEE
- [12] YueZhang, James B.D. Joshi. An Implementation Architecture of the GTRBAC Model (2010). Information Science Department University of Pittsburgh Pittsburgh, P A, US.
- [13] Ramadan Abdunabi, Mustafa Al-Lail, Indrakshi Ray, and Robert B. France Specification, Validation, and Enforcement of a Generalized Spatio-Temporal Role-Based Access Control Model (September 2013).
- [14] Oracle, "Core J2EE Patterns - Data Access Object". Internet access: <http://www.oracle.com/technetwork/java/dataaccesobject-138824.html>
- [15] Network Working Group - Request for Comments:2866, "RADIUS Accounting", Livingston June 2000. Internet access: <https://tools.ietf.org/html/rfc2866>
- [16] Network Working Group - Request for Comments:2838, "Remote Authentication Dial In User Service (RADIUS)", Livingston April 1997. Internet access: <https://tools.ietf.org/html/rfc2138>
- [17] Novell, masterSoft "NTRadPing 1.5 RADIUS Test Utility" Internet access: <https://www.novell.com/coololutions/tools/14377.html>
- [18] JUnit framework. Internet access: <http://junit.org/>
- [19] Binghao Li, James Salter, Andrew G. Dempster and Chris Rizos, "Indoor Positioning Techniques Based on Wireless LAN", School of Computer Science and Engineering, Sydney, Australia.
- [20] Chang N, Rashidzadeh R, Ahmadi M. 2010. Robust indoor positioning using differential Wi-Fi access points. IEEE Transactions on Consumer Electronics.
- [21] Houshang Darabi, Pat Banerjee, Jing Liu, "Survey of Wireless Indoor Positioning Techniques and Systems", IEE Transactions on systems, man, and cybernetics - part C, VOL. 37, NO.6, November 2007
- [22] Reza AW, Geok TK. 2009. Investigation of indoor location sensing via RFID reader network utilizing grid covering algorithm. Wireless Personal Communications.

Finansų rinkų prognozavimas remiantis investuotojų nuotaikų indikatoriumi

Marius Liutvinavičius

Kaunas faculty of humanities,
Vilnius University,
Kaunas, Lithuania
marius.liutvinavicius@khf.vu.lt

Jelena Zubova

Institute of mathematics and informatics,
Vilnius University,
Vilnius, Lithuania
jelena.zubova@mii.vu.lt

Virgilijus Sakalauskas

Kaunas faculty of humanities,
Vilnius University,
Kaunas, Lithuania
virgilijus.sakalauskas@khf.vu.lt

Santrauka — Pastarosios finansų krizės ir besitęsiantis nestabilumas paskatino naujų idėjų paiešką ir intensyvius mokslinius tyrimus finansų rinkų prognozavimo srityje. Naujausi moksliniai tyrimai rodo, kad rinkos funkcionavimas nėra racionalus. Jas veikia ne tik ekonominiai ar socialiniai veiksnių, bet ir investuotojų lūkesčiai, baimės, intuicija, taip pat manipuliavimas žinių srautais bei apskritai informacijos ribotumas. Emocijomis grįstas investavimas lemia nuostolius, o investuotojų nuotaikas atspindintys indikatoriai gali padėti jų išvengti. Šiame tyryme toks indikatorius panaudotas siekiant padidinti investavimo į biržoje kotiruojamą fondą efektyvumą. Ištirtos penkios skirtinges strategijos. Geriausia iš jų leido pasiekti 4,8 % pelningumą, kai tuo pačiu laikotarpiu indikatoriaus nenaudojant buvo patirtas 4 % nuostolis. Rezultatai atskleidžia siūlomų strategijų pranašumą bei investuotojų nuotaikas atspindinčių indikatorių pritaikymo galimybes finansų rinkų prognozavimui.

Raktiniai žodžiai — akcijų rinkos, anomalijų prognozavimas, investuotojų lūkesčiai, dideli duomenys

I. ĮVADAS

Nestabilumas finansų rinkose ir krizinių situacių gausa paskatino naujų idėjų paiešką ir intensyvius mokslinius tyrimus finansų sistemų analizės srityje. Tarptautinės finansų organizacijos ir mokslo astovai atkreipė ypatingą dėmesį į pastarojo meto finansinio sektorius nesaugumą, klasikinių ekonomikos teorijų ribotumą, pabrėžė finansinės informacijos analizės ir prognozavimo mechanizmų tobulinimo svarbą.

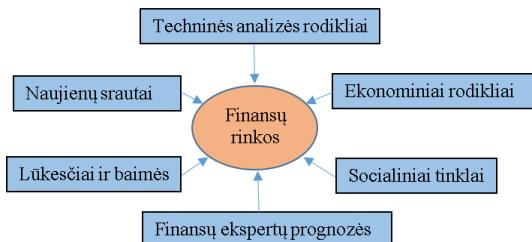
Nusistovėjusios ekonominės teorijos, pabrėžiančios rinkų efektyvumą neatitinka šių dienų realijų. Rinkos kaina gali būti neteisinga dėl gaunamos klaudingos informacijos (viena iš 2008 metų krizės priežasčių – suteiktai ir užtikrinant reitingų bevertėms obligacijoms). Naujausi moksliniai tyrimai įrodo, kad rinkos funkcionavimas nėra racionalus. Jas veikia ne tik ekonominiai, geopolitiniai, socialiniai veiksnių, bet ir investuotojų lūkesčiai, baimės, intuicija, taip pat manipuliavimas žinių srautais bei apskritai informacijos ribotumas.

Visi šie veiksnių yra skirtini, tačiau veikia tiek finansų rinkų judėjimą, tiek realios ekonominės procesus, todėl turi būti įvertinti prognozuojant galimas krizes.

Finansų rinkos elgesys yra dinaminis procesas, kuriame didesnio stabilumo periodus keičia nervingi momentai, kurių metu ženkliai sumažėja rinkos efektyvumas ir padidėja krizinių situacijų tikimybė. Todėl finansų rinkų efektyvumo kitimo stebėjimas – labai svarbus užduavinys, reikalaujantis integruotų techninės ir fundamentaliosios analizės irankių, naujų kiekybinių vertinimo rodiklių bei efektyvių vertinimo metodų. Tai salygoja intelektinių modelių kūrimo poreikių, metodų integravimo, modifikavimo, specializavimo finansinėms rinkoms būtinumą.

Investavimo sprendimai daromi įvertinant daug faktų, bet dažnai ne suvokiama, kaip prendimo rezultatą įtakoja daug papildomų veiksnių, kuriuos būtina suprasti ir įvertinti (naujienu srautai, politiniai interesai, socialiniai pokyčiai, investuotojų lūkesčiai). Rinkų burbulai formuoja dėl šio neteisingo ir nekokybiško informacijos srauto. Būtina sukurti tokius rinkų prognozavimo modelius, kuriais būtų galima patikrinti įtarimus, jog formuoja kainų burbulas, ir nustatyti kurie veiksnių jų lemia.

2013 metų Nobelio ekonomikos premija paskirta R. Shiller [1] (kartu su E. Fama ir L. P. Hansen), tyrusi finansinių aktyvų kainų kitimą ir pabrėžusiam rinkų iracionalumą. Pastaroji finansų krizė leido kitaip pažvelgti į ekonominius procesus, paneigti klasikinį požiūrį į savaimė susireguliuojančias ir racionalias rinkas. Tai atveria kelius naujiems tyrimams, o informacinių technologijų pažanga įgalina šiuos tyrimus. Norint modeliuoti ir integruoti tarpusavyje skirtinges veiksnius, būtina apdoroti su jais susijusius didelius duomenų kiekius, pavyzdžiu naujienu srautus socialiniuose tinkluose.



1 pav. Finansų rinkas įtakojantys veiksnių

Vadovaujamės požiūriu, kad finansų rinkos yra dinaminės tarpusavyje sąveikaujančių agentų sistemos, kurių modeliavimas ir prognozavimas remiasi kompiuterinėmis technologijomis. Šioje srityje tyrimai vykdomi jau daugiau nei dešimtmetį (S. E. Page [17], W. H. Chan i r K. L. Judd [18]), tačiau šiuo metu įgauna naujų formų. Atlikdami finansų rinkų prognozavimo tyrimus remiamės elgsenos ekonomikos (angl. *behaviour economics*) požiūriu (G. Caginalp, V. Smith ir D. Porter [13]). Šiuo metu galime stebėti išaugusį užklausų „bear market“ ir „market crash“ kiekį paieškos sistemose (2 pav.). Pastarajį kartą panaši situacija buvo fiksuojama prieš 2008 m etę finansų krizę. Elgsenos ekonomikos požiūriu tai gali signalizuoti apie gresiančius neramumus finansų rinkose. Tokio t ipo i informaciją galima panaudoti tikslėsiams r izikos lygio įvertinimui.

Bendras mūsų tyrimų tikslas yra ištirti iracionalių veiksmų įtaką finansų krizių formavimuisi. S iekiamo įvertinti rinkos naujienu, socialiniuose tinkluose skelbiamos informacijos bei apskritai investuotojų lūkesčių ir baimių įtaką finansų rinkų stabilumui. Visi šie faktoriai turėtų būti integruoti į rinkos prognozavimo modelius. Šiame darbe tiriama, ar investuotojų nuotaikas atspindis indikatorius gali padėti prognozuoti akcijų kainų kritimus, sumažinti riziką ir padidinti pelningumą.

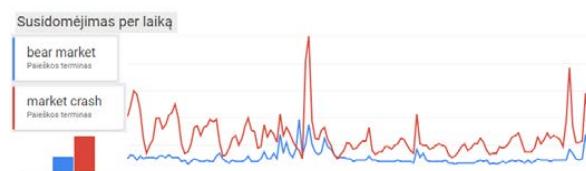
Antrame skyriuje pateikiama finansų krizių prognozavimo modelių apžvalga. Trečiame skyriuje analizuojami prognozavimui naudojami indikatoriai. K etvirtame s kyriuje aprašoma t yrimo audojama metodika. Penktame s kyriuje pateikiami tyrimo rezultatai

II. ANOMALINIŲ SITUACIJŲ PROGNOZAVIMAS

Šioje dalyje pateikiama trumpa finansų krizių prognozavimo modelių apžvalga.

Finansų krizės paprastai apibrežiamos kaip dideli vertybinių popierių kainų kritimai ar platus maisto finansinių institucijų žlugimas [4]. S. Claessens i r M. Kose (2013) pažymėjo, kad yra salyginai nesudėtinga kurti kiekybinius metodus valiutų krizių prognozavimui, tačiau bankų ir skolos krizių prognozavimas paprastai remiasi kokybine ir analitikų nuomone grįsta analize [8].

Pirmosios kartos valiutų krizių prognozavimo modeliai remėsi fundamentaliais ekonomikos rodikliais ir spekuliacijų tikimybe. Pagrindinis dėmesys buvo skirtinas ilgo laikotarpio pusiausvyros t ašķi nustatymui, fiskaliniam deficitui ir monetarinei politikai. Šie modeliai valiutų krizes aiškinė prastomis vidaus makroekonominėmis sąlygomis bei pabréžė ryšį tarp spekuliacinių atakų užsienio valiutų biržose ir makroekonominii rodiklių. [4]



Šaltinis: Google Trends
2 pav. Investuotojų nuotaikų tendencijos

Antrosios kartos modeliai buvo papildyti savaime išspūdžių lūkesčių komponente. Buvo daroma prielaida, kad investuotojų lūkesčiai gali išsauktis valiutų krizes [3].

Trečiosios kartos modeliai apjungė žemus ekonominius rodiklius su prastaais bankinio sektorius rodikliais. Dėl šios priežasties trečios kartos modeliai dar vadinami „dvigubos krizės“: bankų ir valiutų krizės modeliais. [7]

G. Tularam ir B. S. ubramanian (2013) išskyrė 4 krizių prognozavimo modelių kartas. Pasak jų Krugmano (2001) pasiūlytas modelis gali būti priskirtas ketvirtai kartai. Tai jau ne vien valiutų krizių prognozavimo, bet bendresnis modelis, kuriam vertybiniai popierių kainos vaidina pagrindinę rolę. Šios grupės modeliai įvertina ekonominį ir finansinių reguliavimą, akcinių teises, sistemas skaidrumą ir priežiūros lygi, politinius ir sociologinius faktorius [4]. Naujausi ankstyvo krizių prognozavimo modeliai remiasi dideliu kiekiu indikatorių, skirtų matuoti pažeidžiamumą viešajame, finansų, žmonių bei namų ūkių sektoriuose [8].

Finansų rinkos yra sudėtingos sistemas, kuriose didelę reikšmę vaidina žmonių elgsenos ypatumai [4]. Svarbu paminėti, kad agentinėmis sistemomis grįsti modeliai ir simuliacijos gali žymiai efektyviau atskleisti netiesinius elgsenos šablonus nei tradiciniai modeliai. Tai leidžia geriau suvokti investuotojų interesus finansų rinkų funkcionavimo principus bei priežastinius ryšius.

Ankstyvo krizių prognozavimo modeliai remiasi iš esmės dvimi metodais: ekonometriniu ir „signalų“. Ekonometrinio metodo atveju skaičiuojama diskretnaus įvykio – šiuo a tveju krizės pasirodymo arba ne – tikimybė (nuo 0 iki 1). Taip pat jie leidžia nustatyti, kurie nepriklausomi kintamieji turi didžiausią reikšmę priklausomam kintamajam. „Signalų“ metodo atveju naudojamas neparametrinis metodas įvertinti krizės pasirodymo ižizkai. Laiškoma, kad tam tikras faktorius signaluoja galimą krizę, jei peržengia nustatytą ribą. [3]

III. INDIKATORIAI FINANSŲ RINKŲ PROGNOZAVIMUI

A. Ekonominiai ir finansiniai indikatoriai

Ankstyvo finansų krizių prognozavimo modeliai turi daug indikatorių. Kaminsky et al. (1998) atliko tokijų indikatorių apžvalgą ir išskyrė keletą pagrindinių grupių: išorinis sektorius (pvz., einamosios sąskaitos dydis), finansų sektorius, realusis sektorius (pvz., realaus BVP augimas, nedarbo rodikliai), viešieji finansai (fiskaliniai rodikliai), instituciniai kintamieji, politinės situacijos kintamieji.

V. Bucevska (2011) taip pat panaudojo platų indikatorių spektrą: fiskalinius kintamuosius, finansinio sektorius kintamuosius (pvz., bankų išduotų paskolų ir BVP santykis, bankų valdomo turto ir BVP santykis); išorinio sektorius (einamosios sąskaitos) kintamieji (pvz., einamosios sąskaitos deficitas, eksporto augimo tempas); vidinio realaus sektorius kintamieji (pvz., realaus BVP augimo rodiklis).

R. Mariano (2002) spekuliacinių atakų identifikavimui pasiūlė naudoti tokius indikatoriai: laipsniškas tarptautinių rezervų mažėjimas, augantis biudžeto ir einamosios sąskaitos deficitas, laipsniškas valiutų kursų pervertinimas.

G. T ularam ir B. Subramanian (2013) pažymėjo, kad svarbu įvertinti ne tik įvairių rodiklių įtaką prognozuojamam kintamajam, bet ir jų įtaką vienas kitam.

Prognozavimui naudojami įvairūs metodai. G. C . Shen ir W. Y. Jia (2014) i īrdam i K inijos gamybos kompanijų akcijų kritimus naudojo principinių komponenčių metodą pradinės svarbios i informacijos išskyrimui. V ėliau anomalinių situacijų prognozavimo modelis buvo kuriamas panaudojant atraminių vektorių metodą. R. Mariano (2002) pritaikė Markovo grandinių modelį. Liu ir Lindholm (2006) panaudojo neraiškių aibų metodiką nustatant krizę nulemiančius ekonominius veiksnius.

B. Investuotojų nuotaikų indikatoriai

Šiame s kyriuje a p ūvelgiami investuotojų nuotaikas atspindintys indikatoriai.

Elgsenos e konomika (angl. *behavioral economics*) te igia, jog emocijos gali paveikti sprendimų priėmimo procesą. Todėl investuotojų nuotaikas matuojantys indikatoriai gali būti naudingi prognozuojant anomalines situacijas finansų rinkose.

Socialiniai tinklai yra naujas efektyvus informacijos šaltinis investuotojų elgsenos tyrimui ir šablönü nustatymui. Naujausi tyrimai atskleidžia, jog socialinių tinklų duomenys gali padėti prognozuoti finansų rinkas.

Dauguma darbų koncentruojasi į kitos dienos konkretių akcijų kainos prognozavimą. Mes orientuojamės į ilgesnio periodo laikotarpius ir siekiame pritaikyti metodus finansų krizių prognozavimui. Ankstyvas perdėtų lūkesčių suformuotų rinkos „burbulų“ identifikavimas gali leisti išvengti didelių akcijų kainų kritimų.

Pagal G. R anco et. a l. (2015) yra t rys p agrindiniai informacijos šaltiniai: naujienos internete, paieškos sistemų užklausos, socialinių tinklų informacija. Apibendrinant ga lima teigi, kad viena dalis autorų koncentruojasi į žinučių kiekių fiksavimą, o kita į jų turinio analizę. Tiriamas akcijų kainų priklausomybė nuo bendrų bei finansinių naujienų.

P. Houlihan ir G. G. Creame (2014) panaudojo sentimentų, išgautų iš socialinių tinklų informacijos srauto, indikatorių atiesės kainų prognozavimui ir nustatė, kad naujienu srautai ir investuotojų elgesio ypatumai gali paaikiinti akcijų kainų svyravimus.

G. R anco et. a l. (2015) tyrė ryšį tarp Twitter tinklo ir finansų rinkų. Buvo nustatyta stipri korrelacija tarp žinučių kieko ir v ėliau sekantių akcijų kainų svyravimų. Nustatytu, jog nuotaikų socialiniame tinkle pasikeitimais gali prognozuoti akcijų rinkų krypties pasikeitimą. Bollen et al. (2010) pritaikė nuotaikų indikatorių Dow J ones I ndustrial A verage i indeksą prognozavimui.

Kitas būdas investuotojų nuotaikų nustatymui yra paieškos sistemų užklausų analizė. Paprastas tokio tipo analizės pavyzdys pateiktas 2 paveikslė. Išsamus vyraujančių tendencijų nustatymas gali būti naudingas prognozuojant akcijų rinkas. Šioje srityje reikšmingų rezultatų pavyko gauti I. Bordino (2012) ir L. Kristoufek (2013).

Naujienu srautų ir informacijos socialiniuose tinkluose analizė gali atrodyti paprastas ir patikimas dalykas, tačiau svarbu įvertinti tą faktorių, jog dalis žmonių gali užsiimti manipuliacija informacijos srautais. Sukuriant didelius šališkos informacijos srautus gali būti stengiamasi paveikti investuotojų lūkesčius. Svarbu turėti tokius modelius, kurie gebėtų identifikuoti ir eliminuoti klaidinančią informaciją.

Bet kuriuo atveju tyrimai rodo, kad rinkos dalyvių lūkesčiai turi didelę reikšmę vertybinių popierių kainų pokyčiams, todėl tai a tispindintys indikatoriai turėtų būti įtraukti į rinkos prognozavimo modelius.

Mes orientuojamės į anomalinių situacijų prognozavimą. Tradiciškai naudojame anksčiau šiame skyriuje paminėtus ekonominius ir finansinius indikatorius. Tačiau siūlome ši spektrą praplėsti sentimentų indikatoriais (3 pav.).

Dalis tradicinių techninės analizės indikatorių gali būti panaudoti rinkos tendencijoms įvertinti (pvz. *RSI*, *MACD*). Tačiau galima išskirti labiau specifinius indikatorius, atspindinčius investuotojų baimę (pvz., *Volatility index (VIX)* ar lūkesčius (pvz., *Put/Call ratio*, *Bullish percent index*, *Decision Point Volume Trend Oscillator*).

CNN Money Fear and Greed Index yra investuotojų nuotaikas atspindintis indikatorius, susidedantis iš 7 įvairių aspektus įvertinančių rodiklių.

Taip pat jau yra techninės galimybės analizuoti itin didelius socialinių tinklų ir naujienu portalų duomenų srautus bei sukurti tuo paremtus indikatorius. *StockTwits sentiment index* ir *Market Prophit social media sentiment index* įvertina investuotojų lūkesčius pagal jų veiklą socialiniuose tinkluose.

IV. TYRIMO METODIKA

Ankstesni mūsų tyrimai parodė, kad didžiausius nuostolius finansų krizių metu nulėmia neracionalus emocijomis grjistas investuotojų elgesys. Kainoms sparčiai kylančių skubama pirkti, o joms pasiekus dugnā vertybinių popieriai parduodami. Būtent perdėtas optimizmas gali vertybinių popierių kainas iškelti aukščiau jų realios vertės, o perdėta baimė pernelyg nusmukdyti. Mes darome prialaidą, kad sentimentų indikatorių panaudojimas gali padėti eliminuoti neracionalų elgesį ir išvengti nuostolių.

Paprastai šie indikatoriai naudojami p rognouojant k itos dienos konkretios įmonės akcijų kainą. Mes siekiame šiuos indikatorius p ritaikyti ilg ola ikotarpio investavimui ir anomalinių situacijų prognozavimui.

Tradiciniai techninės analizės indikatoriai	Lūkesčių ir baimės indikatoriai	Socialinių tinklų sentimentų indikatoriai
<ul style="list-style-type: none"> • MACD • Average directional index (ADX) • Put/Call ratio • RSI 	<ul style="list-style-type: none"> • CNN Money Fear and greed index • ISE sentiment index • Decision Point Volume Point Ascillator (VTO) • Volatility index (VIX) • Bullish percent index 	<ul style="list-style-type: none"> • Stock twits sentiment index • Market prophit social media sentiment index • HedgeChatter social sentiment indicators

3 pav. Nuotaikų indikatoriai

Tyrimui naudojamas indikatorius, analogiškas CNN Money Fear and Greed Index [2]. Jis susideda iš:

- VIX indeks;
- Indekso S&P 500 vertės palyginus su 125 dienų slenkančio vidurkio reikšme;
- santykio tarp Niujorko biržos akcijų, kurių kainą viršija 52 savaičių aukštumas, ir tų, kurios pasiekė žemumas;
- put/call santykio;
- santykio tarp kylančių ir krintančių akcijų;
- skirtumo tarp akcijų ir obligacijų pelningumo;
- skirtumo tarp investicinių reitingų turinčių obligacijų grąžos ir „šiukslinių obligacijų“ grąžos.

Nuotaikų indikatoriaus įtaka tiriamą imituojant periodinį investavimą į ETF fondą. Periodiškai už nustatytą pinigų sumą yra perkami ETF fondo vienetai. Nupirktu vienetų kiekis priklauso nuo tos dienos vieneto kainos. Taip pat nuo vieneto kainos priklauso ir viso portfelio vertė. Sprendimai dėl įmokos dydžio ir ar apskritai verta p īrkti/parduoti fondo v ienetus priklauso nuo nuotaikų (sentimentų) indikatoriaus (SI) reikšmės. Jeigu ji didelė, reiškia vyrauja optimizmas, o jeigu maža, tuomet dominuoja baimė.

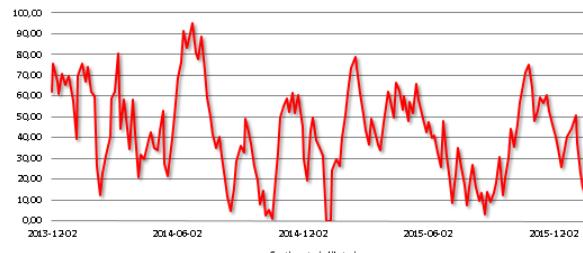
Visų pirmą atsižvelgiant į šį indikatorių yra nustatomas įmokos dydis (5 pav.). Kai indikatoriaus reikšmė yra ypatingai žema ir peržengia nustatyta apatinę (Ra) ribą, tuomet įmoka yra didinama pradinę reikšmę padauginant iš tam tikro koeficiente (Kd). Darome prialaidą, kad vyraujant didelei baimei, akcijų kainos būna žemiau savo realios vertės, todėl tai gera proga investuoti. Jeigu indikatoriaus reikšmė yra labai didelė ir peržengia nustatyta viršutinę ribą (Rv), tuomet įmoka yra mažinama dauginant iš koeficiente Km, kuris gali igyti reikšmes nuo 0 iki 1. Tokiu būdu stengiamasi užfiksuoti situacijas, kai lūkesčiai yra per dideli, ir išvengti pervertintų vertybinių popierių pirkimo.

Už nustatytą pinigų sumą yra perkami fondo vienetai:

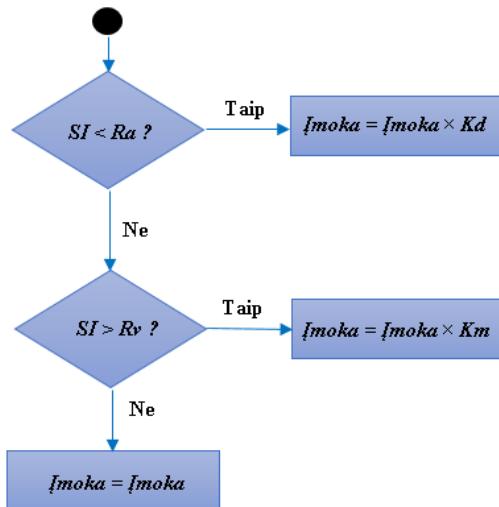
$$Nupirkti\ vienetai = \bar{I}moka / Vieneto\ vertė$$

Naujai nupirkti vienetai pridedami prie anksciai nupirkto vienetų, taip didinant bendrą investiciniame portfelelyje esantį vieneto kiekį. Identifikavus didelę riziką, sukaupti vienetai yra parduodami už tos dienos vieneto kainą ir pašalinami iš investicinio portfelio:

$$Sukaupti\ vienetai_t = Sukaupti\ vienetai_{t-1} + Nupirkti\ vienetai_t - Parduoti\ vienetai_t$$



4 pav. Nuotaikų indikatoriaus reikšmės



5 pav. Įmokos dydžio nustatymo schema

Portfelio vertė priklauso nuo sukauptų vienetų kieko ir vienetu kainos konkrečiu metu:

$$Portfelio\ vertė = Sukaupti\ vienetai \times Vieneto\ kaina$$

Jeigu i ndikatoriaus reikšmės labai i šauga i r peržengia nustatyta ribą Rs, tuomet visi sukaupti fondo vienetai yra parduodami. Taip siekiama išvengti staigū ir didelių portfelio vertės sumažėjimų. Tokiu momentu gaunamos investavimo pajamos, apskaičiuojamos pagal formulę:

$$Investavimo\ pajamos = Parduoti\ vienetai \times Vieneto\ kaina$$

Kiekvieną kartą pardavus fondo vienetus gautos pajamos yra kaupiamos. Pasibaigus investavimo laikotarpiui pelnas yra apskaičiuojamas pagal formulę:

$$Pelnas = Visos\ investavimo\ pajamos - Visos\ įmokos$$

Mokesčių įtaka nėra vertinama. Tiriama, ar indikatoriaus naudojimas gali padidinti pelningumą.

V. REZULTATAI

Siekiant ivertinti investuotojų nuotaikų indikatoriaus reikšmę priimant sprendimus buvo imituojamas investavimo į biržoje kotiruojamą fondą (SPDR S&P 500 [12]) procesas. Skaičiavimams naudotos šio fondo istorinės vieneto vertės. Investavimo laikotarpis – 2 metai (nuo 2014 m. sausio mėn. iki 2016 m. sausio mėn.). Šis periodas pasirinktas dėl jo specifiko ir panašumo į krizinę situaciją, k uomet p atiriami nuostoliai – 20 mėnesių nuolatos kilęs indeksas per kitus 4 mėnesius patyrė 2 korekcijas ir gržo prie savo pradinį reikšmių. Standartinis periodinės įmokos dydis – 100 EUR. Konkrečius perkamu ir paduodamų fondo vienetu kiekiai nustatomi pagal 4 skyriuje aprašytą metodiką.

Iš viso analizuotos 5 skirtinės strategijos, kurios naudoja investuotojų nuotaikų indikatorių. Jų parametrų reikšmės pateikiamas 1 lentelėje. Šie rezultatai palyginti su investavimu nenaudojant indikatoriaus – t. y. tiesiog sekant indeksą.

I LENTELĖ. STRATEGIJŲ PARAMETRAI

	Rs	Kd	Ra	Km	Rv
Strategija Nr. 1	60	50	5	0	55
Strategija Nr. 2	50	10	10	0,5	60
Strategija Nr. 3	65	20	10	0,2	65
Strategija Nr. 4	70	30	10	1	70
Strategija Nr. 5	55	40	5	0,1	60

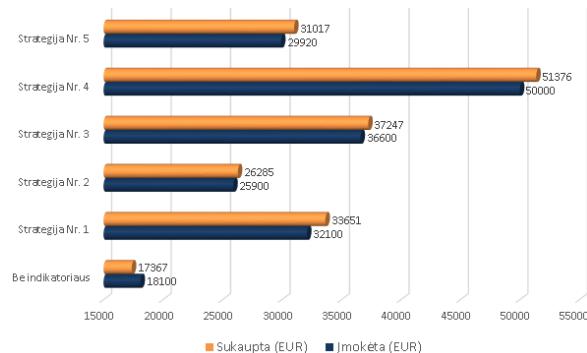
Rezultatai rodo, k ad t i piškas investavimas s ekant indeksą analizuojamu laikotarpiu būtų lémęs nuostoli – būtų sukaupta mažiau pinigų nei iš viso įmokėta (6 pav.). Tuo tarpu visais penkiais atvejais, kai buvo naudojamas nuotaikų indikatorius, nuostolių išvengti pavyko. Dél naudojamų skirtingų koeficientų ženklai skyrësi investuotos sumos.

Daugiausia i nvestuota i r ka rtu daugiausia s ukaupta b uvo naudojant strategiją Nr. 4. Šiuo atveju sukaupti vienetai buvo parduodami tik jei indikatoriaus reikšmë bûdavo didesnë už 70. Įmokos d ydis ne buvo m azinamas ne t jei reikšmës viršydavo viršutinę ribą. Tuo tarpu indikatoriaus reikšmëms nukritus žemiau 10, įmokos buvo padidinamos 30 kartų.

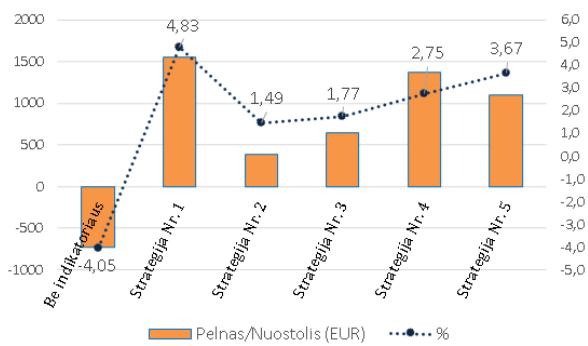
Strategija N r. 2 buvo daug konservatyvesnë, todél investuotos ir sukauptos sumos daug mažesnës. Šiuo atveju fondo vienetai b uvo p arduodami v os indikatoriaus reikšmë pasiekdavo 50. Jeigu j i vi ršydavo 60 , tuomet įmokos dydis buvo mažinamas per pusę. Reikšmei nukritus žemiau 10 įmoka buvo padidinama 10 kartų.

Vis délto svarbiausias dalykas yra skirtingų strategijų pelningumas. Šiuo atveju strategija Nr. 4 ne tik atnešë daugiau pelno nei strategija Nr. 2 (atitinkamai 1 376 EUR ir 385 EUR), bet ir buvo santykinių pelningesnë (atitinkamai 2,75 % i r 1,49 %).

Taip pat svarbu palyginti strategijas Nr. 4 ir Nr. 5. Pastaroji nors ir davë mažiau g ryojo pelno (1 097 EUR), tačiau jos pelningumas siekë 3,67 %. Mažesnë įmoketą ir sukauptą sumą lémë konservatyviau apskaičiuojami įmokų dydžiai bei dažniau paimamas pelnas. Tačiau rizikos ir pelningumo rodikliai šiuo atveju buvo geresni.



6 pav. Strategijų rezultatų palyginimas



7 pav. Pelningumo palyginimas

Vis délto geriausius rezultatus davë strategija Nr. 1. Ji maksimaliai išnaudojo investuotojų nuotaikų indikatorių. Jo reikšmei pakilus vi rš 60 vi s ukaupti vi enetai buvo parduodami, taip siekiant išvengti nuostolių. Reikšmei nukritus iki 5, įmoka buvo didinama 50 kartų, siekiant maksimaliai pasinaudoti galimai nepagrįstai nukritusia kaina. Tuo tarpu indikatoriaus reikšmei viršijus 55 įmoka buvo mažinama iki 0. 7 pav. rodo, jog ši strategija leido pasiekti 4,83 % pelningumą. Bendras pe lnas (1551 E UR) taip pa t buvo di džiausias. Palyginus šios strategijos rezultatus su i nvestavimu nenaudojant nuotaikų indikatoriaus galime matyti daugiau nei 8 % geresnį rezultatą. Rezultatai rodo, jog nuotaikų indikatoriaus naudojimas gali padëti išvengti didelių portfelio vertës sumažėjimų esant akcijų kainų korekcijoms bei pasinaudoti palankiomis progomis, kai rinkose vyrauja baimë ir aktyvų kainos pernelyg nuvertintos.

VI. IŠVADOS

Emocijomis gristas investavimas lemia nuostolius, o investuotojų nuotaikas atspindinčių indikatorių naudojimas gali padëti jų išvengti. Šiame tyime nuotaikų indikatorius buvo pritaikytas periodiniam investavimui į biržoje kotiruojamą fondą. I nvestuojamų sumų dydžiai ir pinigų atitraukimo iš rinkos momentai buvo apskaičiuojami remiantis indikatoriaus reikšmëmis.

Užfiksavus pernelyg didelį optimizmo lygį buvo siekiama išvengti galimai pervertintų aktyvų pirkimo, todél įmokos buvo mažinamos. Didelés indikatoriaus reikšmës taip pat laikyti signalu, jog gali išvysti kainų korekcija, todél tokiai atvejais sukaupti vienetai buvo parduodami, taip siekiant išsaugoti sukauptą pelną. Situacijos, kai i ndikatorius rodë ypač didelį pesimizmą, laikyti palankiomis progomis investuoti daugiau. Tokie ve iksmai iš esmës yra priešingi tipiškam investuotojų elgesiui, dél kurio patiriama nuostoliai.

Ištirtos penkios skirtingos nuotaikų indikatoriaus naudojimu paremtos strategijos. Visais atvejais rezultatai buvo geresni nei i t uomet, ka i ndikatorius ne buvo naudojamas. Geriausia strategija, m aksimaliai iš naudojanti indikatoriaus galimybës, leido pasiekti 4,83 % pelningumą, kai tuo pačiu laikotarpiu i ndikatoriaus n enaudojant buvo pa tirtas 4 % nuostolis. Gauti rezultatai parodo, jog investuotojų nuotaikas atspindys indikatoriai gali būti pritaikyti finansų rinkų prognozavimui bei rizikos valdymui.

LITERATŪRA

- [1] R. S hiller, „Speculative Asset Prices”, American Economic Review, 2014, 104(6): 1486-1517. Available: <http://cowles.yale.edu/sites/default/files/files/pub/d19/d1936.pdf>
- [2] Fear & Greed Index. Available: <http://money.cnn.com/data/fear-and-greed/>
- [3] V. Bucevska, „An analysis of financial crisis by an early warning system model: The case of the EU candidate countries”, BEH - Business and Economic Horizons, Volume 4, Issue 1, January 2011, p. 13-26. ISSN: 1804-1205.
- [4] G. A. Tularam, B. Subramanian, „Modeling of financial crises: a critical analysis of models leading to the global financial crisis”. Global journal of business research, Volume 7, nr. 3, 2013.
- [5] G. C. Shen, W. Y. Jia, „The Prediction Model of Financial Crisis Based on the Combination of Principle Component Analysis and Support Vector Machine”, Open Journal of Social Sciences, 2, 204-212, 2014. Available: <http://dx.doi.org/10.4236/jss.2014.29035>
- [6] R. S. Mariano, B. N. Gültekin, S. Ozmucur, „Models of Economic and Financial Crises”, Proceedings of the Middle East Economic Association, 2002, ISSN: 2334-282X.
- [7] T. R. Singh, „An ordered probit model of an early warning system for predicting financial crisis in India”, Proceedings of the IFC Conference on „Initiatives to address data gaps revealed by the financial crisis”, Basel, 2010.
- [8] S. Claessens, M. A. Kose, „Financial Crises: Explanations, Types, and Implications”, IMF Working Paper, 2013.
- [9] P. Houlihan, G. Creame, „Can social media and the options market predict the stock market behavior?”, Stevens Institute of Technology, 2014. Available: https://editoralexpress.com/cgi-bin/conference/download.cgi?db_name=CEF2015&paper_id=521
- [10] G. Rancho, D. Aleksovski, G. Caldarelli, M. Grčar, I. Mozetič, „The Effects of Twitter Sentiment on Stock Price Returns”, 2015, PLoS ONE 10(9): e0138441, doi:10.1371/journal.pone.0138441.
- [11] H. Schoen, D. Gayo-Avello et. al., „The power of prediction with social media”, Internet Research, 2013, Volume 23, Issue 5, p. 528-543, 10.1108/IntR-06-2013-0115.
- [12] State Street Global Advisors, „SPDR S&P500 ETF”. Available: <https://www.spdrs.com/product/fund.seam?ticker=SPY>
- [13] G. Caginalp, V. A. Ilieva, D. Porter, V. L. Smith, „Do Speculative Stocks Lower Prices and Increase Volatility of Value Stocks?”, The Journal of Psychology & Financial Markets, Volume 3, 2002. Available: <http://ssrn.com/abstract=312421>
- [14] J. Bollen, H. Mao, X. Zeng, „Twitter Mood Predicts the Stock Market”, 2010. Available: <http://arxiv.org/abs/1010.3003>
- [15] I. Bordino, S. Battiston, G. Caldarelli, M. Cristelli, A. Ukkonen, I. Weber, „Web search queries can predict stock market volumes”, Plos One, 2012, 7 (7): e40014.
- [16] L. Kristoufek, „Can Google Trends Search Queries Contribute to Risk Diversification?”, Scientific reports, 2013, 3. doi: 10.1038/srep02713.
- [17] S. E. Page, „Computational models from A to Z”, Complexity, 5: 35-41, 1999. doi: 10.1002/(SICI)1099-0526(199909/10)5:1<35::AID-CPLX5>3.0.CO;2-B
- [18] W. J. Den Haan, L. K. Judd, M. Juillard, „Computational suite of models with heterogeneous agents II: Multi-country real business cycle models”, Journal of Economic Dynamics and Control, 2011, Elsevier, Volume 35(2), p. 175-177.

SUMMARY

The continued economic instability demands for new ideas and methods for financial markets forecasting. Financial markets are influenced not only by economic or social factors. Expectations and fears of investors, the impact of media streams and social networks must be incorporated in financial crisis prediction models. In this work we present how the indicator of investors' mood can reduce the risk and increase profitability. Five strategies of such indicator based investment were investigated. All of them outperformed the common investors' behaviour. They allowed to avoid losses and reduced the volatility of portfolio.

Comparison of popular PHP frameworks

Ernestas Burokas

Kaunas University of Technology, Faculty of Informatics
Kaunas, Lithuania
ernestas.burokas@ktu.edu

Dominykas Barisas

Kaunas University of Technology, Faculty of Informatics
Kaunas, Lithuania
dominykas.barisas@ktu.lt

PHP frameworks are meant to facilitate software development and are becoming more and more popular in web based projects. Before starting the project it is extremely important to choose framework that goes along well with the system and customer requirements. The decision is impeded by a wide variety of frameworks available in the market. A considerable effort is required for the user to get acquainted with all the popular frameworks, configure environments, prepare test cases and evaluate framework effectiveness and other characteristics.

In order to facilitate the evaluation process, a tool for PHP framework comparison was developed. Research based on this tool was conducted using six of the most popular frameworks (Symfony2, CakePHP, Zend2, Laravel, CodeIgniter and Yii). Results were compared in four characteristics: speed, class usage, code usage and cyclomatic complexity.

Keywords—PHP; frameworks; Laravel; Symfony2; CakePHP; Zend; CodeIgniter; Yii; comparison

I. INTRODUCTION

Decision to analyze PHP programming language frameworks was made because of lack of literature about most popular PHP frameworks and due to the framework popularity. Main purpose of this research is to pick out the best framework based on main characteristics. To achieve this goal a web tool was created. This tool allows comparing frameworks using different characteristics with different amount of data and structure in database. To make research more extensive, a plain PHP (no framework) comparison was included. It is beneficial to evaluate the cases when PHP frameworks might not be necessary.

II. POPULARITY OF PHP FRAMEWORKS

It is fair to say that PHP is leading programming language in web site development in these days. According to statistics [1], 81.8% of all sites are written in PHP, in the second place with 15.9% is .NET. These two languages are taking 97.7% of all available websites.

Looking at all programming languages frameworks popularity among the websites (Fig. 1) it is obvious that PHP frameworks play a major role. Diagram shows that about 40 percent of websites are using PHP programming language frameworks. Next after PHP is .Net which have 21.84% of frameworks usage. Finally, other less known languages are taking small part of website usage [2].

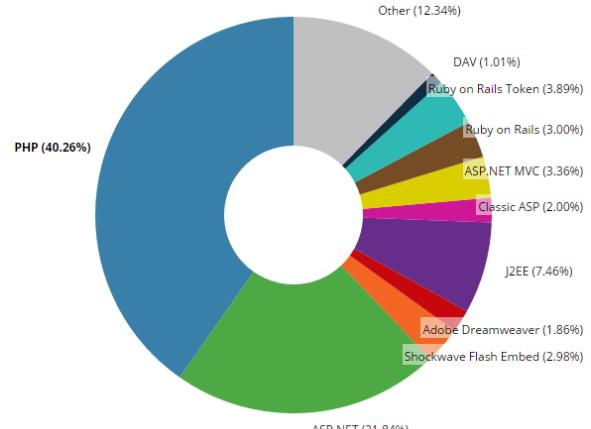


Fig. 1. Frameworks usage in websites

III. ARCHITECTURE OF PHP FRAMEWORKS

Programming language framework is collection of libraries used to increase system maintainability and reduce time of coding. PHP programming language frameworks have a lot of built-in functions such as: access to database, session management, security etc. These functions help to reduce coding and testing effort for developer, because it eliminates the need of manual implementation of various features. Most of the frameworks also maintain simple and clear folder structure. It keeps different framework parts apart from each other and facilitates file manipulation [3] [2].

All frameworks that are addressed in this research are using MVC model. The model isolates user interface, business logic and data model from each other and defines relationships between them. Well-designed MVC project enables independent work between front-end and back-end developers with no interaction.

MVC pattern consist of three major parts: model, view and controller. Data and relationship between different parts are shown in Fig. 2. First of all, user sends request to the controller, where the request is analyzed. Further, the controller calls model responsible for the logic and connection with database to get data that is needed to perform request. When model finishes job all data is returned back to the controller. Finally, all data is returned to view and page is rendered [4].

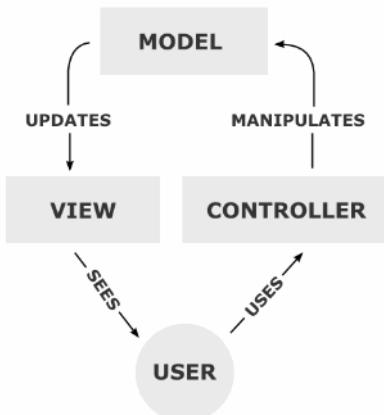


Fig. 2. MVC pattern

Model is like connection between controller and view. Essentially model is sort of “blind”, because it is not capable knowing what happens with all data, when the data is passed to another component. The main purpose of model is to prepare data and pass it to the controller. However, model differs from database. Model works like data guard, which accepts all requests and processes them. In most cases this part of MVC pattern is the most difficult compared to other parts [4] [5].

View is the result of rendered data, which comes from the model. Usually, view is a part of HTML. For example, button in a webpage is generated by view. When user clicks the button controller action is triggered. Most common mistake done by developers is the absence of connection between view and model. In that case all data comes from the controller which breaks all principals of the MVC pattern [4].

Last part of MVC pattern is controller. Main task for this part is to process data which comes from the user and accordingly update model. Controller is the only part, which interacts with the user. Without user actions controller would not make sense. Controller is like data assembler, which pass data to the model. Controller can have only logic which appoints to data collection. Besides, controller is connected to one view and model only to ensure one directional data passing.

IV. RELATED WORK

In 2013 research to compare two PHP frameworks has been conducted. During the research three different versions of the same system was created. First version was in plain PHP (with no framework), other two was developed in CakePHP and CodeIgniter. Those systems was created using Eclipse PDT IDE. All versions were published in local WAMP server. Applications were compared in three different characteristics: code lines number, speed and security [2].

Total lines of code were determined using CLOC (Count Lines of Code) software. Looking at the results (table I), authors claim that using plain PHP code is not reusable and all needed libraries should be coded by the developer. So it means that without using any framework it requires to write much

more code. Frameworks are using a lot of libraries with built-in functions, which makes code cleaner. Comparing CakePHP with CodeIgniter it is seen that CakePHP requires less coding, because it has built-in system, which connects database to the model. Writing less code means faster system development, better maintainability and improved security [2].

TABLE I. TOTAL CODE LINES OF EACH APPLICATION

	PHP	CSS	PHP files	CSS files
Plain PHP	3002	650	42	13
CodeIgniter	2321	621	36	13
CakePHP	1408	680	16	13

To test applications speed XDebug was used. This tool provides functionality like function trace, memory distribution and time to perform tasks. Speed results are introduced in table II. From the results it can be stated that plain PHP has huge advantage compared to frameworks. For example to complete login/logout operation for plain PHP code it takes only 96 milliseconds, while using CodeIgniter 2291 milliseconds. Authors claim that CodeIgniter is faster than CakePHP because most of the operations are executing faster with CodeIgniter. Frameworks are working slower because they need to load a lot of libraries while executing operations and it takes time.

TABLE II. SPEED COMPARISON OF EACH APPLICATION

Functionality	Plain PHP (ms)	CodeIgniter (ms)	CakePHP (ms)
Login/Log-Out	96	2291	1559
Session management	91	739	1309
Database	143	762	1492
Forms usage	487	1338	2262
Forms validation	321	966	2411

To compare applications security, Nessus tool was selected. According to results CakePHP has better security system than CodeIgniter and plain PHP. CakePHP framework has built-in security components, which grants protection from main web vulnerabilities. While CodeIgniter framework has similar protection system (CSRF and XSS protection), but during the tests it does not protect from SQL injection attack.

Similar research was arranged in Bleking Technology Institute, Sweden in 2012 by student. Same frameworks (CakePHP, CodeIgniter) were compared [6]. Main purpose of this analysis was to compare speed of PHP frameworks. After all, six tests were released, three of them were using database. There were 20000 requests completed to Apache server. During the tests (using database) frameworks selection (in each request) was done randomly. Results are shown in table III. According to results CakePHP is a little bit faster (by 5.34 milliseconds) than CodeIgniter [6].

TABLE III. TESTS USING DATABASE

	Response times (ms)	
	CodeIgniter	CakePHP
1 test	227	216
2 test	215	211
3 test	209	206
Average	217	211.6

Other three tests were executed without database (Table IV). Results show that response time using frameworks is almost the same. CakePHP was slightly faster (1.3 millisecond) than CodeIgniter. So in both tests results are very similar, CakePHP is a little bit faster than CodeIgniter framework.

TABLE IV. TESTS WITHOUT DATABASE

	Response times (ms)	
	CodeIgniter	CakePHP
1 test	233	233
2 test	230	229
3 test	229	226
Average	230.66	229.33

Comparing all six frameworks on their available functionality, table V with results are presented. Results reveal that ORM is available on all frameworks except Zend. Authentication is not implemented only in CodeIgniter framework. Popular functions like cache and validation is available in all compared frameworks. All frameworks supports PHPUnit testing tool, but Yii framework also use Selenium for testing. Looking at the other features, Symfony2 have advantage, because it has layers and CRUD generation components. Furthermore, together with Laravel, these frameworks support logging functionality. Talking about security, only Zend and Yii frameworks does not have protection against web vulnerabilities. In general Symfony2 and Laravel frameworks have most of the features. In the other hand Zend framework is in the last place [7].

One way to compare PHP programming language frameworks is their popularity. For that purpose Google Trends tool was used [8]. This tool can show how many requests by keyword was in search engine. According the results oldest framework is Zend, because in 2004 first requests with this keyword appeared. Besides that Zend framework had reached biggest popularity among all frameworks in the end of 2009. But from that time this framework popularity was going down, so now it is the last framework by this characteristic (16% of best framework search). Fifth place is taken by CakePHP framework, which searching numbers starting to grow in middle of 2005 and in 2010 reaches maximum, now there is 18% of popularity among frameworks. Yii framework is in fourth place with 21% of popularity. This framework is quite new, had an increasing popularity since 2009 and maximum

was reached in 2014. Third place between PHP frameworks goes to by Symfony2 which has 30% of popularity. Symfony2 framework started to grow in 2006 and in 2010 reached maximum, from that day it has stable popularity among users. CodeIgniter with 50% is in second place. The most popular framework of the search engine is Laravel. This framework is strong leader of this characteristic. Besides that it is newest framework (started to grow in 2012).

TABLE V. FRAMEWORK FEATURES

	Cake PHP	Yii	Zend 2	Symf. 2	Larav.	CodeI.
Latest version	3.2	2.0	2.5.2	3.0	5.2	3.0.4
ORM	+	+	-	+	+	+
Authorization	+	+	+	+	+	-
Cache storage	+	+	+	+	+	+
Dependency injection	-	-	+	+	+	-
Code generation	+	+	-	+	+	-
Testing library	+	+	+	+	+	+
Logging management	-	-	-	+	+	-
XSS	+	-	-	+	+	+
XSRF	+	-	-	+	+	+
SQL injection	+	-	-	-	+	+
Template system	+	+	-	+	+	+
Services	+	+	-	-	+	+
CRUD Generation	-	-	-	+	-	-

V. METHODOLOGY OF THE RESEARCH

To compare PHP language frameworks web-based tool was created. It was developed using PHP programming language with no framework. In general, this web application make request to pages that are created using PHP framework and collects the results. All results are calculated (taking average) and compared in one graph for each characteristic. The comparison tool has a feature to add custom data table with custom data. This data can also be generated automatically. After user has completed the customization and preparation of his own data, PHP framework comparison tool starts sending subsequent requests and collecting the result data. During this research custom data table was created. Data table consist of five columns: first name (string), last name (string), age (integer), birthday (date), married (bool). To maintain realistic situation about 300000 records was inserted into table.

Another seven applications were created for PHP frameworks (one for plain PHP). It is simple one page applications which render page with special hidden inputs with data needed for comparing those frameworks. Those inputs have page characteristic data like loaded classes, code lines and cyclomatic complexity. Each framework has the same

functionality. It gets data from database and renders that data as table in page. Of course, each PHP framework code is different, because of a functions available for a particular framework. For example, some of the frameworks supports query builder for SQL, others don't. All applications are hosted on the local XAMPP server. Local environment has following characteristics: Windows 10 education OS, 16 gigabytes of RAM, 2.5 gigahertz quad core CPU. To complete all functionality two databases were used. One for main application (compares PHP frameworks), which store user information also PHP frameworks information used for analysis. Other database is for remaining applications, it is storing demo user data. All frameworks are compared by four characteristics.

First one is speed. This characteristic was selected because of importance of fast loading pages these days. Performance of page loading is essential characteristic, because users don't want to wait too long. Speed characteristic is used to define how fast page is rendered using one of PHP programming language framework. To get this information main application is initiating JavaScript Ajax call to a page. Time intervals between request and response are measured. Usually, to make good comparison of PHP frameworks a significant amount of measurements (in this case, that amount was 10) need to be collected to avoid accidental results. All response times are added and average is taken.

Next characteristic is an amount of loaded classes needed to render the page. This characteristic purpose is to define how many classes from PHP framework libraries are used to complete various kinds of actions. This characteristic is highly related with speed, because using more classes for the page rendering, means longer time to load all files. Moreover, number of classes could give some insights to a scale of functionality available to the user.

Other characteristic is an amount of code lines per page. It's very useful characteristic especially for developers, because it's good to know how many code lines is required to perform some functionality. The purpose of this characteristic is to define total amount of PHP code lines written to perform operation, in this case to render the page. Code lines also show a code complexity, because larger amount of lines mean higher complexity and more complex maintainability.

Last characteristic is a cyclomatic complexity. This characteristic is related to all different paths that can be executed within a function. This characteristic is very handy to define complexity of the code. Besides, other conclusions or recommendations can be derived from this metric. For example, greater cyclomatic complexity means harder testing of application.

VI. RESULTS

Speed characteristic measured for a custom applications based on all frameworks is shown in table VI. Results are provided in five different data record sizes, where one record size corresponds the amount loaded in one page and inserted into table. Results indicate that the fastest method to render data can be achieved without a framework. Despite this, the best results in terms of speed were achieved by Laravel and Yii

frameworks. While CakePHP took the longest period in time to perform the same tasks. Other frameworks had similar results.

TABLE VI. SPEED PER REQUEST ON EACH APPLICATION MEASURED IN MILLISECONDS

Name	Data amount				
	100	1000	10000	50000	100000
Laravel	0.136	0.138	0.353	1.988	4.982
CakePHP	0.196	0.281	1.374	7.94	19.626
Symfony2	0.289	0.394	0.711	2.889	6.435
CodeIgniter	0.036	0.054	0.272	1.986	5.138
Yii	0.062	0.072	0.264	1.913	4.967
Zend2	0.106	0.121	0.364	2.204	5.603
Plain PHP	0.017	0.031	0.186	0.949	1.282

Judging from the results at 100, 1000 and 10000 data records all frameworks has quite similar results. But when data is growing to 100000 records CakePHP starts working really slow (fig. 3). So in general, most stable results were presented by Laravel and Yii frameworks and of course plain PHP has best speed characteristics.

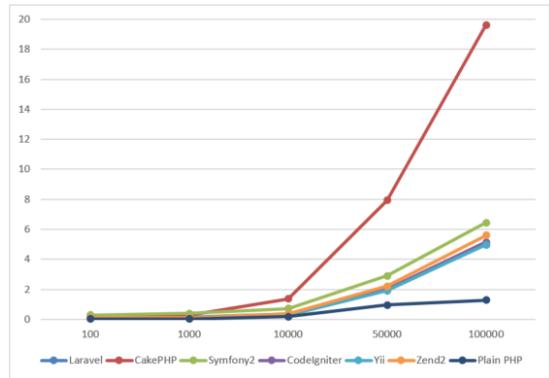


Fig. 3. Dependency between speed characteristic (in milliseconds) and amount of data

Comparison of other three characteristics of PHP frameworks are shown in table VII. According to the results, the most lines of code are used by plain PHP and Zend2 frameworks. Leaders in this characteristic with five lines of code are Laravel and CakePHP. In case of the measurement of loaded classes, it is clear that the largest amount of classes needed for Symfony2 framework. Plain PHP is loading default PHP classes only and therefore requires a small amount of classes to be loaded.

TABLE VII. PHP FRAMEWORK CHARACTERISTIC COMPARISON

Name	Code lines	Loaded classes	Cyclomatic complexity
Laravel	5	302	2
CakePHP	5	270	2
Symfony2	7	438	2
CodeIgniter	6	176	2
Yii	13	212	2
Zend2	18	323	4
Plain PHP	23	159	4

Most of the frameworks have similar value for cyclomatic complexity that means code simplicity in that applications. This is good thing when it comes to maintainability and testing. Zend2 framework and plain PHP have a bit higher cyclomatic complexity than other frameworks.

CONCLUSIONS

After analyzing the results collected from the local server, the following conclusions can be drawn:

1. In terms of popularity and the richness of features Laravel can be distinguished as a leading framework.
2. Experimental evaluation (speed, loaded classes, lines of code and cyclomatic complexity) gave different values to various features of a distinct frameworks. An in-depth study of these results need to be performed in order to choose the most suitable framework for a particular situation. In general, Laravel outperformed other frameworks and appeared to a leader in a majority of the characteristics.
3. Theoretical comparison of the frameworks is very similar to practical research results. So it is fair to say that most popular framework by people search on Google is also best in other characteristics.

4. As expected, without using any framework has advantages such as faster request processing, but on the downside making application without framework requires better programming skills, effort and security understanding since there is no built-in security functions.

There are several additional features to be added to the framework comparison tool in the future:

1. Creating code-blocks for each framework with a similar functionality and the same naming convention while trying to find the most effective solution for each framework.
2. Introducing additional metrics and stress testing with a large number of concurrent connections for each application created with different framework.

REFERENCES

- [1] World Wide Web Technology Surveys, online access: <http://w3techs.com/>
- [2] L. Lancor and S. Katha, "Analyzing PHP frameworks for use in a project-based software engineering cours cte", Proceeding of the 44th ACM technical symposium on Computer science education, 2013.
- [3] W. Cui, L. Huang, L. J. Liang, J. Li, "The Research of PHP Development Framework Based on MVC Pattern", Conference on Computer Sciences and Convergence Information Technology, IEEE Computer Society, 2009.
- [4] C. Hopkins, "The MVC pattern and PHP", 2013.
- [5] C. Supaartagorn, "PHP Framework for database management based on MVC pattern", Department of Mathematics Statistics and Computer, 2011.
- [6] H.Nylen, „PHP Framework Performance for Web“, Computer Sciences and Convergence Information Technology, 2012.
- [7] „PHP frameworks comparison“,online access: <http://socialcompare.com/en/comparison/php-frameworks-comparison>
- [8] Google Trends, online access: <https://www.google.com/trends/>

Finansinių piramidžių modeliavimas

Ridas Ramonas

Vilniaus universiteto Kauno humanitarinis fakultetas
Kaunas, Lietuva

Dr. Vera Moskaliova

Vilniaus universiteto Kauno humanitarinis fakultetas
Kaunas, Lietuva

Finansinės piramidės (FP) yra gana senas ir ekonomikoje gerai žinomas reiškinys, galintis sukelti įvairių ekonominį, socialinių bei kitokių neigiamų padarinijų, kurių dažnai gali būti labai sunku išvengti, ypač jei investuojama su skolintu kapitalu. Dėl šios priežasties yra svarbu žinoti tokį schemų veikimo principus, nustatyti susidarymo priežastis bei ieškoti prevencijos būdų. Straipsnyje trumpai pristatoma FP samprata bei aptariami egzistuojančios FP tyrimo požiūriai. Taikant logistinę kapitalo valdymo teoriją, skaičiuojamos pasirinktos investicinės bendrovės sukauptojo kapitalo esamoji bei būsimoji vertės, tiriamos, kokioms sąlygoms esant ši investicinė bendrovė taptų finansine piramide.

Raktiniai žodžiai: finansinės piramidės, Ponzi schemas, modeliavimas, logistinė teorija, kapitalo valdymas.

I. FINANSINIŲ PIRAMIDŽIŲ SAMPRATA

Terminas „finansinė piramidė“ yra gana senas ir daugeliui žmonių suprantamas, nes ne vienas yra susiviliojęs lengvai uždrabamas pinigais ir taip nukentėjęs. Istorijoje šiuo metu galima aptikti labai daug FP pavyzdžių, pavyzdžiui, C. Ponzi 1920 m. sukurta afera ar dar palyginus visai nesenai (2008 m.) išaiškinta B. Madoff afera. Šis reiškinys labai išplito po 1990 m. (ypač 1994 m.), kai suklestejo FP daugelyje skirtinguoju pereinamojo periodo ekonomiką [1]. Šiuolaikinės FP yra pateikiamos išradinėja ir daugiausia veikia internete, todėl šio reiškinio tyrimas yra būtinas.

Plačiau išnagrinėjus FP tyruisių mokslininkų darbus ([2], [3], [4], [5], [6], [7]) bei juose pateikiamus apibréžimus, galima sakyti, jog FP gali būti apibūdinta kaip sukčiavimo priemonė, kurios metu investuotojams pasiūlomas tariamas investicinės produktas, turintis labai mažą rizikos laipsnį bei ypač aukštą investicijų grąžą. Nepaisant to, jokios investicijos nėra įvykdamos, dėl to šiai priemonei yra būtinas vis kitų (naujų) investuotojų pritraukimas, kad būtų išmokėti pelnai jau esamiems investuotojams, o FP galėtų ir toliau egzistuoti. Būtent ši požymij galima nurodyti kaip vieną pagrindinių, dėl ko visos FP yra pasmerktos žlugti, nes visada egzistuoja galutinis investuotojų skaičius.

II. FINANSINIŲ PIRAMIDŽIŲ TYRIMO POŽIŪRIAIS

Tariant mokslinę literatūrą, gali būti pastebeti įvairūs skirtini FP bei kitų nestabilių situacijų tyrimų požiūriai. [11] darbe išskiriama 5 požiūriai, kuriais gali būti nagrinėjamos FP. Prie šių požiūrių dar galima pridėti H. Minsky išplėtota pesimistiškų įvykių vertinimą [8]. Dėl to iš viso išskiriama 6 FP tyrimo požiūriai:

A. Pesimistiškas požiūris

Šis požiūris yra aptartas H. Minsky finansinio nestabilumo teorijoje 1960 ir 1970 m. Pasak autoriaus, finansinėje sistemoje stabilumas yra paradoksaliai destabilizuojantis: „geri laikai“ skatina eksperimentuoti bei prisimti per didelę riziką, o tai gali baigtis su trenksmu. Ši teorija priveda prie dviejų išvadų: 1) finansinė sistema iš prigimties yra nestabilė ir ji turi būti visada prižiūrima; 2) nestabilumas negali atitikti vieno modelio, o tiksliai žlugimo prigimtis ir rezultatas priklauso nuo konkrečios šalies ekonomikos charakteristikų [9].

B. Racionalusis požiūris

Antrojo tyrimo požiūrio šalininkai, O. J. Blanchard ir M. Watson, remiasi racionalumo prielaida ir teigia, jog visų rinkos dalyvių elgesys yra racionalus. Remiantis visuotinio racionalumo teorija, matyti, kad FP gali egzistuoti. Jos, kaip ir finansiniai burbulai, gali būti aiškinamos kaip labai greitas kainų augimas, kurio su turima informacija paaiškinti neįmanoma, nes jis nesuderinamas su racionalumu. Šis kainų augimas gali augti tik iki tam tikro lygio, o po to kainos staigiai mažeja [10].

C. Žaidimo požiūris

Trečiosios teorijos šalininkai A. Belianin ir O. G. Issoupova FP laiko žaidimu, žvelgdami iš žaidimų teorijos pusės, tačiau neatsieja jos nuo individualių investuotojų azartinų lošimų. Žaidimo dalyviai gali būti visi norintys bei turintys tam lėšų gyventojai. Šio žaidimo piramidinės schemas organizatorius siekia pritraukti kuo daugiau žaidėjų, taip didindamas savo pelną [11].

D. Scenarijaus požiūris

Ketvirtuojo požiūrio autorius S. V. Dubovski teigia, jog FP žlugimas yra apibréžtas jos ypatybėmis. Autorius daro prielaidą, jog piramidinės sistemas organizatoriaus sukaupti finansiniai išipareigojimai yra padengiami tik naujai surinktomis lėšomis, neperžengiant FP ribų. Egzistuojant šiai prielaidai, yra savaike suprantama, jog FP ikturėjas tam tikru laiko momentu gali nebeturėti užtektinai lėšų padengti savo finansinius išipareigojimus, o tai veda FP žlugimo link [12].

E. Įsiskolinimų išipareigojimų sistemos požiūris

Penktuojo požiūrio atveju FP tiriamas kaip per laiką auganti įsiskolinimų išipareigojimų sistema, kuri yra palaikoma tik laiko atžvilgiu didėjančiomis įnešamomis investicijomis. Ši FP galėtų egzistuoti ilgai, jeigu išipareigojimų kiekis neaugtų per greitai ir santykis su pajamomis išliktu ganėtinai pastovus.

Šiame darbe daroma išvada, jog kuo FP įkūrėjui bus didesnis noras pritraukti lėšų, tuo greičiau ateis FP žlugimas [11].

F. Logistinis požiūris

Šeštojo požiūrio atveju, nagrinėjant FP, turėtų būti atsižvelgiama į rinkos prisotinimą, nes realiomis sąlygomis, ypač tiriant uždarą ekonomiką, kapitalo augimas vienodu tempu negali vykti labai ilgai. Žvelgiant iš logistinio požiūrio pusės, kapitalo augimas pats kuria sau konkurenciją, todėl dar sutikta konkurencija iš išorės lėtina jo augimą, kol jis visai sustoja, o sistema prisotinama kapitalu. Svarbi logistinio požiūrio charakteristika yra našumas (arba „talpa“), tai yra, populiacija turi tam tikrą prisotinimo lygį, kurį pasiekus kapitalas pradeda nykti [11].

Toliau šiame straipsnyje ir bus remiamasi būtent logistiniu požiūriu, kuris plačiai aptariamas dagelyje darbų ([11], [13], [14], [15], [16]). Nagrinėtuose darbuose pateikiamos šios teorijos pritaikymo prielaidos bei atliekami tyrimai nagrinėjant FP susidarymą tiek per esamąją, tiek per būsimąją sukauptojo kapitalo vertę, kai investuojant į FP yra naudojanamas nuosavas kapitalas.

III. FINANSINIŲ PIRAMIDŽIŲ BŪSIMOS SUKAUPTOJO KAPITALO VERTĖS TYRIMAS

Būsimos sukauptojo kapitalo vertės tyrimui būtų galima pasinaudoti S. Girdzijausko siūloma logistine būsimos sukauptojo kapitalo vertės skaičiavimo lygtimi [15]:

$$S_n = \sum_{m=1}^{\infty} S_m = a \sum_{m=1}^{\infty} (1 - mq^m) \cdot q^{m-n} e^{-qm} \quad (1)$$

Šioje lygyje S_n – sukauptosios pinigų sumos būsimoji vertė, m – pinigų srauto periodo eilės numeris, n – kaupimo trukmė mėnesiais, a – mėnesis pradžioje įmokama pinigų suma, q – palūkanų procentas nuo pradinės sumos, q – jnašų kiekių didėjimo ar mažėjimo koeficientas, r – investicinės jnašų dydžio augimo koeficientas su palūkanų norma ($r = 1 + i$).

Lygyje daroma prielaida, jog FP veikia kaip ir investicinė bendrovė, o tai reiškia, jog gautą kapitalą kartu investuoja su mėnesio palūkanų norma [15].

Sukauptojo kapitalo būsimosios vertės skaičiavimui ypač didelę reikšmę daro 3 kintamieji: mokamas palūkanų procentas, skaičiuojant nuo pradinės sumos (q), jnašų kiekių kitimo koeficientas (q) bei investicinės jnašų dydžio augimo koeficientas su palūkanų norma (r). S. Girdzijauskas ir V. Moskaliovą po atliktu tyrimu nustatė, jog kuo didesnė q tuo FP tampa nestabilesnė. Vertinant jnašų kiekių kitimą buvo nustatyta, jog kuo jnašų kiekių kitimo koeficientas yra mažesnis, tuo FP stabiliu veikia. Tiesa, jnašų dydžio augimo koeficiente didėjimas FP daro stabilesnę, nes gaunama didesnė investicinė grąža iš FP jnašų [11].

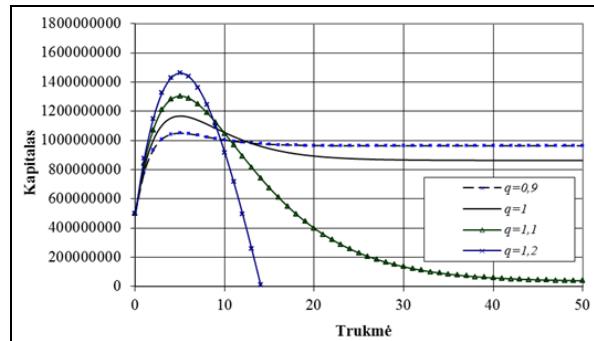
Tiriant sukauptojo kapitalo būsimąją vertę yra pasinaudota šiuo metu internetinėje erdvėje *Poker Automatics* siūlomu pasyvius uždarbio planu. Norint pasinaudoti šiuo pasiūlymu, tereikia investuoti konkrečią sumą pinigų ir iš to bus gaunamas

pasyvus uždarbis kiekvieną mėnesį. Kaip teigama šios investicinės bendrovės internetiniame puslapyje, pajamos yra gaunamos iš sukurtos automatizuotos sistemos, kuri žaidžia pokerį internete.

Sistemos siūlomas uždarbis priklauso nuo investuotos sumos dydžio, pavyzdžiui, investavus 1 000 eurų per mėnesį, siūloma užsidirbtī 18 % (depozitą galima atsiimti po 2 mėnesių), o investavus 50 000 eurų – net 24 % palūkanų per mėnesį (depozitą galima atsiimti po 6 mėnesių). Si investicinė bendrovė siūlo daug investicinių pasirinkimų, tačiau bendra uždirbama suma iš investicijų visada yra apie 30 %. Beje, kuo mažesnė investuojama suma, tuo didesnė dalį investuotos sumos pasiliake pati investicinė bendrovė.

Pasinaudojus šios investicinės bendrovės siūlomomis sąlygomis bei (1) sukauptojo kapitalo būsimosios vertės analize, galime ištirti tam tikras situacijas, kurioms susidarius ši investicinė bendrovė virstų FP.

Skaičiavimai atliekami atsižvelgiant į *Poker Automatics* 3 lygio (18 % grąža per mėnesį) siūlomas sąlygas. Taigi β (pastovią išmokamą palūkanų normą) galime prilyginti 18 %, a (jnešamo kapitalo dydį) – 1 000 eurų, n (kaupimo trukmę) – 50 mėnesių, pirmuoju atveju i (investavimo mėnesio palūkanų normą) laikysime esant 30 %, o q (jnašų skaičiaus kitimo koeficientą) keisime nuo 0,9 (nuolat mažėjantis jnašų skaičius) iki 1,2 (nuolat augantis jnašų skaičius). Galime ištirti, kaip esant šioms sąlygoms keičiasi sukauptos sumos būsimoji vertė, priklausomai nuo kaupimo trukmės, esant skirtingam jnašų skaičiaus kitimui.

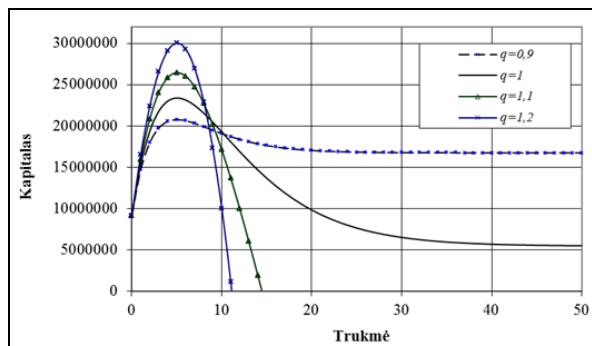


1 pav. Sukauptojo kapitalo būsimosios vertės priklausomybė nuo kaupimo trukmės, kai $a=1000$; $r=1,3$; $n=50$; $\beta=0,18$

Ivertinus šios investicinės bendrovės pateikiamus duomenis galima pirmame paveiksle matyti, jog investicinės bendrovės rezultatai tampa nestabilūs, tik esant didžiausiam jnašų skaičiui, o mažėjant jnašų skaičiui investicinės bendrovės gaunami rezultatai yra stabilūs ir ji negalėtų būti vadina FP. Tiesa, šie rezultatai yra gauti pagal pačios investicinės bendrovės pateikiamus duomenis, kurie nebūtinai gali būti teisingi, todėl toliau būtų galima pasižiūrėti, kaip keistusi kreivės, jeigu visus kintamuosius, išskyrus investavimo mėnesio palūkanų normą, paliktume nepakitusius.

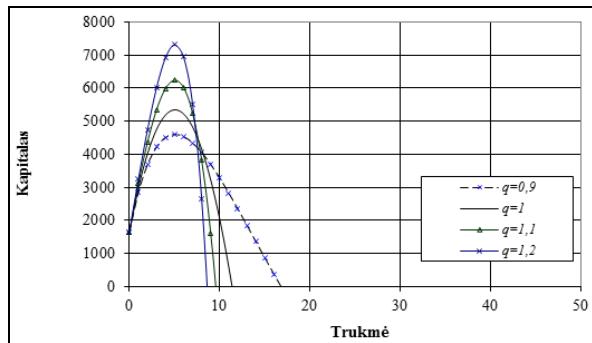
Antrame paveiksle matyti, kad sumažinus investavimo palūkanų normą iki 20 % ir esant tam pačiam, arba nuolat mažėjančiam, jnašų skaičiui, investicinė bendrovė dar gali

stabiliai funkcionuoti. Jeigu jnašų skaičius vis augtų ir nepavyktų surinkto kapitalo investuoti su didesne nei 20 % palūkanų norma, investicinė bendrovė stabilumo neišlaikytų ir palyginus per gana trumpą laiką (~5 periodus) pasiektų maksimaliai įmanomą sukaupti kapitalą, o tada per ateinančius 6–9 periodus patirtų didelių nuostolių, nes sukaupto kapitalo sumaaptu minusinę.



2 pav. Sukauptojo kapitalo būsimosios vertės priklausomybė nuo kaupimo trukmės, kai $a=1000$; $r=1,2$; $n=50$; $\beta=0,18$

Toliau galima pasižiūrėti, kaip keistusi sukauptas kapitalas, jeigu ši investicinė bendrovė neturėtų žadamos investavimo strategijos ir sukauptą kapitalą investuotų tik su 1 % palūkanomis (3 pav.).



3 pav. Sukauptojo kapitalo būsimosios vertės priklausomybė nuo kaupimo trukmės, kai $a=1000$; $r=1,01$; $n=50$; $\beta=0,18$

Trečiame paveiksle matyti, jog visais atvejais ši bendrovė aptaptų FP, kuri, net ir mažėjant jnašų skaičiui, prarastų sukauptą kapitalą jau per 18 periodų.

Apibendrinant gautus rezultatus matyti, jog visais tirtais atvejais, kai β koeficientas buvo lygus 18 %, maksimali sukaupti suma buvo pasiekima per 5 periodus. Atliktas tyrimas atskleidė, kad pagal investicinės bendrovės pateikiamus duomenis ši investicinė bendrovė veiktu stabiliai tol, kol jnašų augimo koeficientas neviršytų 1,1. Žinoma, mažinant investavimo palūkanų normą stabilumo situacija prastėjo, tačiau visais atvejais pasireiškė ta pati tendencija, jog kuo mažesnis jnašų skaičius, tuo stabiliau veikia investicinė bendrovė.

IV. FINANSINIŲ PIRAMIDŽIŲ ESAMOS SUKAUPTOJO KAPITALO VERTĖS TYRIMAS

Esamos sukauptojo kapitalo vertės tyrimui atlkti galima pasinaudoti S. Girdžiausko siūloma logistine esamos sukauptojo kapitalo vertės skaičiavimo lygtimi [15]:

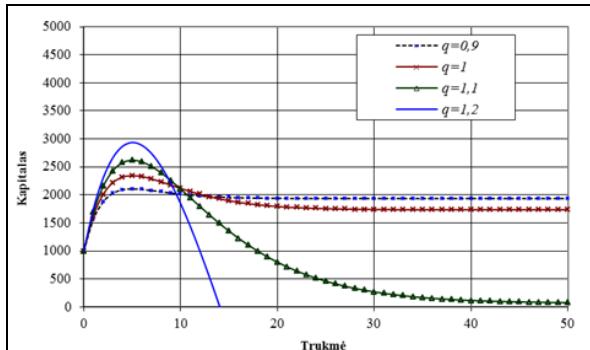
$$S_m = \sum_{n=1}^{\infty} a M_m = \sum_{n=1}^{\infty} a m! \cdot (m+1)^m \cdot \frac{1}{n^{m+1}} \quad (2)$$

Šioje lygtje koeficientų reikšmės išlieka nepasikeitusios, todėl S_m – sukauptosios pinigų sumos esamoji vertė, m – pinigų srauto periodo eilės numeris, n – kaupimo trukmė mėnesiais, a – mėnesio pradžioje įmokama pinigų suma, i – palūkanų procentas nuo pradinės sumos, q – jnašų kiekio didėjimo ar mažėjimo koeficientas, r – investicinės jnašų dydžio augimo koeficientas su palūkanų norma i ($r = 1 + i$).

Ši lygtis yra šiek tiek patogesnė nei (1), nes joje nereikia iš anksto žinoti skaičiavimais įvertintų periodų skaičiaus, kadangi čia apskaičiuojama pinigų srauto dabartinė vertė bet kuriai norimai trukmei [11].

Skaičiuojant kapitalo dydį šia lygtimi yra gaunami labai panašūs rezultatai kaip ir skaičiuojant su būsimosios vertės lygtimi. Beje, šiuo atveju yra gaunami kiek mažesni sukaupti pinigų srautai, nes yra įvertinamas laiko veiksny ir jo įtaka sukaupai pinigų sumai [14].

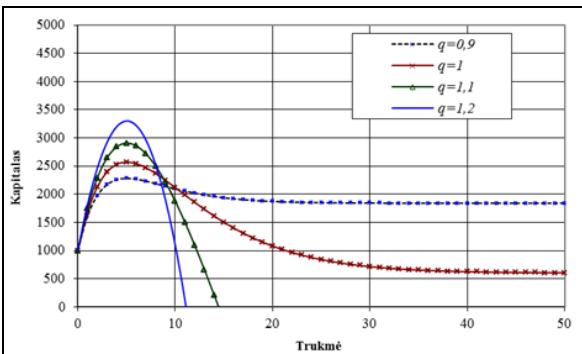
Esamos sukauptojo kapitalo vertės tyrimui atlkti naudojamas tomis pačiomis tiriamos investicinės bendrovės siūlomomis sąlygomis.



4 pav. Sukauptojo kapitalo dabartinės vertės priklausomybė nuo kaupimo trukmės, kai $a=1000$; $r=1,3$; $n=50$; $\beta=0,18$

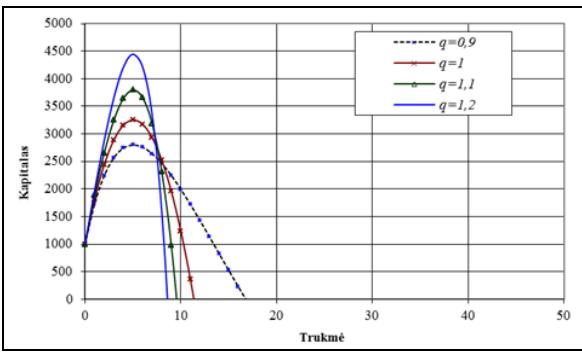
Kaip matyti 4 paveiksle, šiuo atveju yra gauti labai panašūs rezultatai kaip ir 2 paveikslė grafiuke, skiriasi tik gauta sukauptojo kapitalo suma. Investicinė bendrovė nestabiliu veiktu tik tuo atveju, jei jnašų augimo greitis viršytų 1,1, o didžiausias sukauptas kapitalas būtų 5 periode.

Penktu paveikslė rezultatai atskleidžia, kad investicinė bendrovė dar galėtų stabiliai funkcionuoti, jei investavimo palūkanų norma būtų sumažinta iki 20 % ir jnašų skaičius mažėtų arba išslyktų tokis pats.



5 pav. Sukauptojo kapitalo dabartinės vertės priklausomybė nuo kaupimo trukmės, kai $a=1000$; $r=1,2$; $n=50$; $\beta=0,18$

Šeštame paveiksle pateikiami duomenys, kaip ir prieš tai pateiktuose grafikuose, rodančiuose sukauptą dabartinės vertės priklausomybę, atitinka duomenis, gautus tiriant sukauptojo kapitalo būsimosios vertės priklausomybę. Kadangi maksimali kapitalo suma sukaupiama per 5 periodus, net ir su labai maža palūkanų norma investicinė bendrovė visais atvejais išlieka nestabili.



6 pav. Sukauptojo kapitalo dabartinės vertės priklausomybė nuo kaupimo trukmės, kai $a=1000$; $r=1,01$; $n=50$; $\beta=0,18$

Apibendrinant pateikiamus grafikus, galima patvirtinti V. Moskaliovos ir S. Girdzijausko padarytą išvadą, jog stabiliausias investicinės bendrovės atvejis gaunamas tada, kai q koeficientas yra mažesnis už vienetą, arba kai įnašų į investicinę bendrovę skaičius mažėja [11]. Kitaip tarant, bendras FP stabilumas didėja mažejant įnašų skaičiui. Šios investicinės bendrovės atveju ji dar išlikę stabili, jei įmonės investicijų grąža iš gautų įnašų sumažėtų iki 20 % ir įnašų skaičius išlikę tokis pats arba mažėtų. Didinant įnašų skaičių arba mažinanči investicijų grąžą pastebimas greitas kapitalo mažėjimas.

V. IŠVADOS

Galima teigti, jog FP tyrimas bei modeliavimas yra būtini, norint atpažinti FP bei nuo jų apsisaugoti. Iš atlirkų tyrimų matyti, jog net ir normaliai funkcionuojanti investicinė bendrovė, pasikeitus tam tikroms sąlygomis, tam tikru metu galiapti FP.

Ivertinus sukauptojo kapitalo esamos bei būsimos vertės priklausomybę nuo kaupimo trukmės, matyti, jog bendras FP stabilumas didėja mažejant įnašų skaičiui. Tirtos investicinės bendrovės atveju ji išlikę stabili, jei įmonės investavimo palūkanų norma sumažėtų iki 20 % ir įnašų skaičius išlikę tokis pats arba mažėtų. Didinant įnašų skaičių arba mažinanč investicijų grąžą pastebimas greitas sukaupto kapitalo mažėjimas.

Tirta investicinė bendrovė veiktu stabiliu tol, kol įnašų augimo koeficientas neviršytų 1,1. Sukaupto kapitalo vertė mažėja smunkiant investavimo palūkanų normai, tačiau visais atvejais pasireiškia la pati tendencija – kuo mažesnis įnašų skaičius, tuo stabiliu veikia investicinė bendrovė.

LITERATŪRA

- [1] A. Belianin, O. Issoupova, „Financial pyramids in transitional economies. A game-theoretic approach“, Maskva: EERC, 2001, 70 p.
- [2] J. Edward McMillan, „Preventing fraud in nonprofit organizations“, New Jersey: John Wiley & Sons, Inc., 217 p.
- [3] C. L. Pozza, T. R. Cox, R. J. Morald, „A review of recent investor issues in the Madoff, Stanford and Forte Ponzi scheme cases“, Journal of Business & Securities Law, 10 leidinys, 2009, 113–131 p.
- [4] A. B. Garner, „What is ponzi scheme?“, Black's Law Dictionary, 9 leidimas, 2009, 1920 p.
- [5] T. Anderson, H. Lane, M. Fox, „Consequences and responses to the Madoff fraud“, Journal of International Banking and Regulation, 24 (11), 2009, 548–555 p.
- [6] R. Bale, T. Volpe, „Ponzi schemes and financial fraud litigation“, Brief, 40 leidinys, 2011, 8 p.
- [7] T. Frankel, „The ponzi scheme puzzle“, New York: Oxford University Press, 1 leidimas, 2012, 256 p.
- [8] A. Nesvetailova, „Ponzi finance and global liquidity meltdown: lessons from Minsky“, City University London, 2008, 13 p..
- [9] P. Hyman Minsky, „The financial instability hypothesis“, The Jerome Levy Economics Institute of Bard College, 1992, 9 p.
- [10] J. Oliver Blanchard, W. Mark Watson, „Bubbles, rational expectations and financial markets“, National Bureau of Economic Research, 1982, 30 p..
- [11] V. Moskaliova, S. Girdzijauskas, „Finansinių piramidžių stabumo modeliavimas“, Informacijos moksmai, 35 leidinys, 2005, 158–169 p.
- [12] S. V. Dubovskij, „Прогнозирование инфляции и обменного курса рубля в российской нестационарной экономикею“ Москва: Издательство УПСС, 2001, 40 c.
- [13] S. Girdzijauskas „Draudimas: kiekybinė finansinė analizė“, Kaunas: Naujasis lankas, 2002, 104 p.
- [14] S. Girdzijauskas „Finansinė analizė: kiekybiniai metodai“ Vilnius: VU leidykla, 2005, 356 p..
- [15] S. Girdzijauskas „Logistinė kapitalo valdymo teorija; determinuotieji metodai“ Vilnius: VU leidykla, 2006, 232 p.
- [16] V. Moskaliova „Netvarų ekonominė situacijų logistinė analizė“, Daktaro disertacija. Vilnius, VU leidykla, 2009, 156 p.

Oscillator generated signal strength evaluation method

Rokas Mateika

Vilnius University, Faculty of Humanities, Muitines str. 8, LT-44280 Kaunas, Lithuania, EU

Abstract — The main objective of this paper is to present an analysis of a method, when RSI, Momentum oscillators as well as volume and price change signals were combined together to research signal profitability according to signal strength. Combining signals together led to one unique signal which strength was analysed by researching the trading performance and evaluating the profitability of trades. Analysis was held in with different market trends (downtrend, uptrend and flat) of Dow Jones Industrial Average index.

Keywords — Oscillators, RSI, Momentum, Volume, Price change, DJIA, signal strength.

I. INTRODUCTION

Modern day stock trading has gone far beyond financial statement reading, market monitoring and price change graphs understanding. Along with the previous stated, a considerable amount of indicators are used to verify whether the exact moment is the right time to enter the position. Despite the high amount of indicators, using them can significantly improve profitability [1], [2], [3]. Furthermore there are also a lot of publications that consider combining indicators together and modifying them to receive a better result [4], [5], [6]. It should be noted that high amount of indicators and signals can make it hard for the trader to choose one stock over the other, therefore the strength of an indicator signal, as a criteria, has to be considered [7].

Using various kinds of oscillators or in general indicators without signal strength is very subjective and what is more, it is hard to evaluate different stocks, especially in the same market, when the same signals occur. Hard evaluation leads to unevaluated risk to choose the worse investment over a better one. Furthermore, when signal strength is not evaluated, it is hard to test a certain strategy, due to the fact, that indicators occur only certain amount of time on certain stocks before changing. Therefore, using the signal strength evaluation methodology can help indicate the right time for buying stocks, changing stocks in the portfolio or choosing one stock over the other.

A contribution of this paper lies in a RSI and Momentum oscillator and one of volume change and / or price change indicator combination (RMVC), when all of the mentioned indicators signals are joined together and a summarizing signal with different strength levels developed. Further sections of this paper present the basic concept of the proposed methodology, experimental investigations and results.

II. THE BASIC CONCEPT OF THE PROPOSED METHODOLOGY

The proposed strategy is based on combination of RSI and Momentum oscillators, that are furthermore strengthened with trading volume and price change indicators [4]. This strategy is proposed after numerous back-tests when different types of oscillators and their combinations were analysed to get the best average one trade profitability.

As mentioned above, after numerous back-tests final combination of RMVC was settled. RMVC combination was tested with Down Jones Industrial Average (DJIA) stocks, with the historical data from the year of 2008, 2014, 2015. Combination of these three years allows testing the strategy with downtrend and uptrend markets as well as when the market trend is flat (overall change per year is close to zero). After testing data from different years it can be summarized, that RMVC methodology is profitable despite the market trend.

The methodology was purified by going through different combinations of indicators.

Firstly, the performance of separate oscillators (RSI, Momentum, ROC, Stochastic, W%R) was calculated.

Secondly, all the same oscillators were combined with additional trading volume and (or) price change indicators [4], when these indicators went at least 10% above 30 trading days average.

Table I presents the profitability difference between simple oscillators profitability and oscillators combined with trading volume and (or) price change indicators on year 2015.

TABLE I. DIFFERENT OSCILLATORS METHODS PROFITABILITY EVALUATION

	RSI	Momentum	Stochastic	W%R	ROC
<i>Average profitability (%)</i>					
Oscillators	1,74%	2,29%	0,09%	1,85%	0,81%
Oscillators combined with trading volume and (or) price change indicators	2,46%	2,53%	1,16%	2,38%	2,17%
Difference	+0,72%	+0,24%	+1,07%	+0,53%	+1,36%

Thirdly, any combination of the five oscillators with trading volume and price change indicators were tested. Profitability

was calculated when at least 5 of the indicators showed buying signals. This test was held by calculating profitability of holding a stock from 1 to 10 days. As can be seen in Table II, the test showed that 3 days period is the most profitable and accumulated 3,10% average profit per trade.

TABLE II. BEST PERFORMED PERIOD TESTING

Num. of days	Profitability when >= 5 of the indicators showed buying signals
1	0,24%
2	1,51%
3	3,10%
4	3,04%
5	2,47%
6	1,95%
7	1,28%
8	1,72%
9	1,27%
10	1,54%

Fourthly, after getting 3 days period and profitability data for at least five indicators showing buying signal at once, ROC, Stochastic and W%R oscillators were removed from calculations due to high amount of misleading signals. Therefore, RSI, Momentum, trading volume and price change indicators (RMVC) were left for the final calculations, that brought an average of 3,67% trade profitability within 3 days buying and selling period.

After applying the RMVC methodology, each of the four indicators signals were divided into 3 levels [8] [5], where 3 is the highest level and 1 is the lowest. According to the levels, set values were granted in a range 1-3.

A. RSI levels and values

Testing the methodology, a standard RSI formula (1) was used:

$$RSI = 100 - 100 / (1+RS) \quad (1)$$

RS – Average Gain / Average Loss

RSI oscillators buying signal was taken into account only if the value of the signal is lower than 30.

In Table III RSI oscillators signal levels and values can be found.

TABLE III. RSI SIGNAL VALUES

RSI signal value (x)	Signal value
x < 10	3
10 < x < 20	2
20 < x < 30	1
30 < x	0

B. Momentum levels and values

Testing the methodology, a standard Momentum formula (2) was used:

$$MOMENTUM = (CLOSE(i) / CLOSE(i-N)) * 100 \quad (2)$$

CLOSE (i) — close price of the current bar;

CLOSE (i -N) — close price n bars ago.

Due to Momentum formula it is difficult to unambiguously assign signal level values, this is because Momentum signal can vary according to price change. Testing this methodology Momentum oscillator signal levels and values were calculated according to the Table IV.

TABLE IV. MOMENTUM SIGNAL VALUES

Momentum signal value (x)	Signal value
x < 88	3
88 < x < 92	2
92 < x < 96	1
96 < x	0

C. Trading volume and price change levels and values

Trading volume and price change levels and values calculations have been done in couple of stages. First of all trading volume and price change averages of last 30 days of trading were calculated. Afterwards, the calculated average size was compared the current trading volume and price change. Lastly, if the current indicator size is higher than the average, a set value is assigned. Trading volume and price change levels and values are presented in Table V.

TABLE V. C. TRADING VOLUME AND PRICE CHANGE SIGNAL VALUES

Trading volume / price change (x)	Signal value
Current indicator size is higher by 50% than the average	3
Current indicator size is higher by 30% - 50% than the average	2
Current indicator size is higher by 10% - 30% than the average	1
Current indicator size is higher less than 10% than the average	0

III. EXPERIMENTAL INVESTIGATIONS

The profitability of the presented method was tested using historical DJIA index stock data from 2008, 2014 and 2015.

After conducting the experiment with historical data from 2008, a conclusion was made, that depending on the signal strength value, the profitability raises (see Table VI). This is due to the fact that when the signal strength value equals to 1, the average trade profitability is 0,37 %, when signal strength value is 2, the average trade profitability is 2,08% and when

strength value equals 3, the average trade profitability is 2,37%.

It should be noted, that DJIA index during the year of 2008 lost more than 33% of its value as can be seen in Fig. 1.



Fig. 1. DJIA index graph of 2008

TABLE VI. 2008 DJIA INDEX STOCKS RMVC EXPERIMENTAL INVESTIGATIONS

	Indicator signal value		
	1	2	3
Total signals	86	197	27
Profitable	44	107	16
Not profitable	41	90	11
No change	1	0	0
Average profitability	0,37%	2,08%	2,37%

During 2014, when in the most of stock markets a steady growth was visible, DJIA index rose by more than 9% (see Fig. 2). As for the RMVC methodology, with the indicator signal value 2, the average trade profitability was 2,55% (See Table VII).



Fig. 2. DJIA index graph of 2014

TABLE VII. 2014 DJIA INDEX STOCKS RMVC EXPERIMENTAL INVESTIGATIONS

	Indicator signal value		
	1	2	3
Total signals	52	46	N/D
Profitable	29	24	N/D
Not profitable	23	22	N/D
No change	0	0	N/D
Average profitability	0,51%	2,55%	N/D

During 2015 when most of the stock markets did not reach new highs or lows and more or less stayed in the same level (DJIA index downgraded by 1,86%. See Fig. 3) the tested methodology showed the best results among the three tested years. The best result was identified, when the indicator signal value was 3, then the average trade profit equaled to 7,75% (see Table VIII).



Fig. 3. DJIA index graph of 2015

TABLE VIII. 2015 DJIA INDEX STOCKS RMVC EXPERIMENTAL INVESTIGATIONS

	Indicator signal value		
	1	2	3
Total signals	14	137	6
Profitable	7	86	6
Not profitable	7	51	0
No change	0	0	0
Average profitability	0,18%	1,64%	7,75%

IV. EXPERIMENTAL RESULTS

Experimental results have revealed that RMVC methodology, when the stock is sold after three days from buying, is profitable independently from the market trend, although it is more profitable on the flat market. Furthermore, RMVC methodology with different indicator signal values showed that the higher the indicator signal value, the higher the average trade profit. A summarized DJIA index stock RMVC experimental investigations results are presented in Table IX.

TABLE IX. SUMMARIZED DJIA INDEX STOCKS RMVC EXPERIMENTAL INVESTIGATIONS

Year	Indicator signal value		
	1	2	3
2008	0,37%	2,08%	2,37%
2014	0,51%	2,55%	N.D.
2015	0,18%	1,64%	7,75%

V. CONCLUSIONS

The paper reviewed the proposed trading methodology, when a joint group of indicators were analysed and signal strength value was determined. The research was carried out on

2008, 2014 and 2015 historical data of DJIA index stocks. The results clearly point to a conclusion that the higher the indicator signal value, the higher the average trade profitability..

It is clear that further research in developing the most profitable trading methodology by analyzing different combinations and groups of oscillators can be done. By the way, more than one author tried to find the best profitable trading methodology by analysing different oscillators, their groups and various combinations. A lot of them were successful, but most of the time, their success led to high amount of trades, that led to higher risk and higher taxes and commissions. Due to this reason, it is essential to purify the best time, to enter and close the trade. What is more, most of markets stocks move synchronously, therefore picking the right stock to receive the highest profitability is essential.

REFERENCES

- [1] Cheol-Ho Park, Scott H. Irwin. *The Profitability of Technical Trading Rules in US Futures Markets: A Data Snooping Free Test*. USA. University of Illinois. 2005.
- [2] Antony Jackson, Daniel Ladley. *Market Ecologies: The Interaction and Profitability of Technical Trading Strategies*. United Kingdom. University of Leicester. 2013.
- [3] Min Qi, Yangru Wu. *Technical Trading-Rule Profitability, Data Snooping, and Reality Check: Evidence from the Foreign Exchange Market**. Singapore. Singapore Management University. 2005.
- [4] Chandrapala Pathirawasam. *The Relationship Between Trading Volume and Stock Returns*. Journal of Competitiveness. Issue 3/2011.
- [5] Bruce Vanstone, Gavin Finnie. *Combining Technical Analysis and NeuralNetworks in the Australian Stockmarket*. Australia. Bond University, faculty of business. 2006.
- [6] Antony Jackson, Daniel Ladley. *Market Ecologies: The Interaction and Profitability of Technical Trading Strategies*. United Kingdom. University of Leicester. 2013.
- [7] Robert W. Colby. *The encyclopedia of technical market indicators, McFraw-Hill Professional*. ISBN 0-070-12057-9 978-0-070120-57-0. 2003.
- [8] Christos Alexakis, Panayiotis Alexakis, Manolis Xanthakis. An investigation of price – volume intraday patterns under “Bull” and “Bear” market conditions. European research studies, Volume VI, Issue (3-4), pp. 53-66. 2003.

Lithuanian text document processing and clustering

Greta Ciganaitė Faculty of Informatics, VMU Kaunas, Lithuania Email: greta.ciganaite@fc.vdu.lt	Tomas Krilavičius Faculty of Informatics, VMU Baltic Institute of Advanced Technology, BPTI Kaunas, Vilnius, Lithuania Email: t.krilavicius@bpti.lt	Aušra Mackutė-Varoneckienė Faculty of Informatics, VMU Kaunas, Lithuania Email: a.mackute-varoneckiene@if.vdu.lt
---	---	---

Abstract – Clustering is useful method for text documents organization. It is quite well explored for widely known languages, e.g. English, but for highly inflective languages not all methods can be successfully applied. The goal is to find a set of methods which perform well for Lithuanian language. Four models for statistical feature selection and elimination from feature set were proposed. According to the results of F-score, ten feature sets were selected for further cluster analysis. Documents with various number of features were clustered with K-means, spherical K-means and hierarchical clustering methods. The best results according to F-score value were achieved using spherical K-means algorithm.

Keywords – text document clustering, K-means clustering, Spherical k-means clustering, feature selection and elimination, clustering quality evaluation.

I. INTRODUCTION

Increasing amount of information require novel techniques to handle them. Clustering is a technique, that allows finding similar documents and grouping them into previously unknown categories, or clusters. The idea is to assign similar objects to the same group (cluster) while the difference among these clusters are as big as possible [1]. General clustering, as well as documents clustering, is well known technique [2, 3, 1]. However, most of text documents clustering research explores influence of features construction and selection, as well as clustering methods, on clustering performance [2, 1], and only some results are available for other languages, e.g. arabic [4], Lithuanian [5] and Russian [6, 7]. However, in the later papers feature selection is not investigated in depth, while high number of features slows down clustering considerably, and may negatively influence clustering quality. Therefore, the best performing clustering approach (feature construction, feature selection and clustering algorithm) for Lithuanian, and potentially, other highly inflective languages remains unanswered question.

In this paper we propose and investigate several feature selection techniques, in combination with different feature construction methods (*bag-of-words* and *term frequency – inverse document frequency*) and clustering algorithms (*K-means*, *Spherical K-means* and *Hierarchical clustering*). The results

seem promising (see section III), therefore we plan to continue with other highly inflective languages (e.g. Russian).

In secs. II-A and II-B feature construction and selection methods are discussed respectively. Selected clustering algorithms are presented in sec. II-C. Experimental results are discussed in sec. III.

II. DOCUMENTS PREPROCESSING

A. Documents representation

Documents are represented using *bag-of-words* (BOW) [1]. It can be described as a sparse vector of single and whole words frequencies. Length of each vector is equal m and it contains unique words from whole documents corpus.

In some cases frequencies are not the best way to express an importance of terms because frequent and not so significant terms are the reason why documents could be interpreted as very similar. The opposite situation is with rare terms which could appear not so frequent but the meaning of terms is important in the context [8]. The TF-IDF method reduces this effect. The weight of each term is calculated by following formula

$$\omega_{d,t} = tf_{d,t} \cdot idf_{t,D} \quad (1)$$

where t is a term, D is a corpora, $d \in D$ is a document, $tf_{d,t}$ is frequency of term t in document d , $idf_{t,D} = \ln \frac{N}{n}$, here N is a total number of documents d , $n = |\{d \in D | t \in d\}|$ is a number of documents d in which term t occurs.

B. Feature selection

It is very important to select informative features. We use *document frequency (DF)* and *term weight (TW)* criteria for features selection.

Document frequency (DF) is the number of documents where term t occurs:

$$DF_j = |\{d \mid t_j \in d\}|, \quad (j = \overline{1, m}) \quad (2)$$

Then, the bounds of feature selection are:

$$\alpha_- \leq DF_j \leq \alpha_+ \quad (3)$$

Term weight (TW) is a sum of $\omega_{d,t}$ (1):

$$TW_j = \sum_{i=1}^n \omega_{i,j} \quad (4)$$

and feature is selected as informative if

$$TW_j \geq \alpha_{min}. \quad (5)$$

Here, d is a document, t is a term, α_- , α_+ , α_{min} are freely selected bounds, ω is a weight of term j in document i [9, 10].

We propose four models for informative feature selection based on the document representation and features selection methods discussed above.

- 1) $\mathbf{Q}_{tf}^1 \rightarrow \mathbf{DF} \rightarrow \mathbf{Q}_{tf}^2$ model. At first, $n \times m$ size feature frequency matrix is formed and a part of features were eliminated by eq. 3. The final matrix is an $n \times \ell$ feature frequency matrix \mathbf{Q}_{tf}^2 ($\ell \leq m$).
- 2) $\mathbf{Q}_{tf}^1 \rightarrow \mathbf{DF} \rightarrow \mathbf{Q}_{tfidf}^2$ model is similar to the $\mathbf{Q}_{tf}^1 \rightarrow \mathbf{DF} \rightarrow \mathbf{Q}_{tf}^2$ model, but after informative feature set selection using eq. 3, importance of each feature in documents is recalculated using feature weighting, eq. 1.
- 3) $\mathbf{Q}_{tfidf}^1 \rightarrow \mathbf{TW} \rightarrow \mathbf{Q}_{tf}^2$ model. $n \times m$ feature weight matrix is formed from primary \mathbf{Q}_{tf}^1 matrix using eq. 1, then the features are selected by TW criteria. Uninformative features identified by eq. 5, are eliminated from \mathbf{Q}_{tf}^1 matrix. The final feature frequency matrix \mathbf{Q}_{tf}^2 is $n \times \ell$ size, when $\ell \leq m$.
- 4) $\mathbf{Q}_{tfidf}^1 \rightarrow \mathbf{TW} \rightarrow \mathbf{Q}_{tfidf}^2$ model is similar to $\mathbf{Q}_{tfidf}^1 \rightarrow \mathbf{TW} \rightarrow \mathbf{Q}_{tf}^2$ model. The difference is that in the last stage uninformative features which were identified by eq. 5, are eliminated from \mathbf{Q}_{tfidf}^1 . The final feature weight matrix \mathbf{Q}_{tfidf}^2 is $n \times \ell$, when $\ell \leq m$.

C. Clustering algorithms

K-means (KM) is a variation of more general Expectation-Maximization algorithm [1]. The basic idea of KM algorithm is to move objects from one cluster to another while objective function is minimized [1]. The objective function is a sum of the squared error:

$$SSE = \sum_{i=1}^k \sum_{\vec{x} \in C_i} dist_E(\vec{\mu}_i, \vec{x})^2, \quad (6)$$

where the mean of cluster C_i is described as follows $\vec{\mu}_i = \frac{1}{|C_i|} \sum_{\vec{x} \in C_i} \vec{x}$ ($i = 1, k$). It is quite fast and precise, but the problem in documents clustering is that the distance between documents is measured using Euclidean distance, and it is not suitable for text documents clustering because feature matrices are sparse [11].

The main idea of **spherical K-means** (SKM) algorithm is similar to KM algorithm, i.e. objects are transferred from one cluster to another while change of objective function is less than freely selected value ε :

$$|\mathcal{Q}(\{C_j^{(t)}\}_{j=1}^k) - \mathcal{Q}(\{C_j^{(t+1)}\}_{j=1}^k)| \leq \varepsilon, \quad (7)$$

when

$$\mathcal{Q}(\{C_j\}_{j=1}^k) = \sum_{j=1}^k \sum_{\vec{x}_i \in C_j} dist_c(\vec{x}_i, \vec{\mu}_j). \quad (8)$$

The main difference from KM algorithm is that SKM uses cosine distance and operates on vectors that lie on the unit sphere [12].

The main aim of **hierarchical clustering** is creating a hierarchy of clusters. Usually, agglomerative or divisive approaches are used [1]. We will use an agglomerative method, because it is less complex [13]. At first, all objects are interpreted as separate clusters with one object inside. Then two closest clusters are merged into one cluster. It continues while only one cluster is left [1].

The most important part of hierarchical clustering is choosing a distance measure between clusters, sometimes called *linkage*. Commonly four linkage types are used [1, 3]: Single Link, Complete Link, Average Link and Ward. Based on the preliminary experiments (not reported in the paper due to space limitations) the best performance was achieved using Ward linkage, therefore only results with Ward linkage are reported.

In case of Ward measure two clusters are the closest if after merging the SSE measure increases the least:

$$d_w(C_1, C_2) = \frac{||\bar{C}_1 - \bar{C}_2||^2}{\frac{1}{n_{C_1}} + \frac{1}{n_{C_2}}}, \quad (9)$$

where C_1, C_2 are clusters, n_{C_1}, n_{C_2} are number of objects in clusters C_1 and C_2 respectively, \bar{C}_1, \bar{C}_2 are mean of clusters C_1 and C_2 respectively.

We experiment with Euclidean, cosine and Jaccard distances for SSE calculation.

D. Clustering quality evaluation

Quality of clustering results can be evaluated using so called external and internal indices. Internal indices evaluate relation of intra-cluster and inter-cluster distances, and usually are used for detecting number of clusters and so on. External indices use *a priori* information about documents classes, thus, allowing to evaluate effectiveness of clustering algorithms. In this paper we used only external indices. With *a priori* information available, it is possible to calculate a contingency table I, which shows the identity of performed clustering and objects distribution wrt real classes. We use F-score (a

TABLE I
CONTINGENCY TABLE.

	positive	negative
positive (prediction)	tp	fp
negative (prediction)	fn	tn

harmonic mean of precision P and recall R) for evaluating cluster quality [3, 1].

$$F = \frac{2PR}{P+R} \in [0; 1] \quad (10)$$

where $P = \frac{tp}{tp+fp} \in [0; 1]$ (*precision*) and $R = \frac{tp}{tp+tn} \in [0; 1]$ (*recall*). The higher value of index shows the better clustering.

TABLE II
STRUCTURE OF CORPORA.

(No)	Class name	# of documents
(1)	Auto	499
(2)	Energetics	498
(3)	football	500
(4)	Nature	499
(5)	Youth garden	500
(6)	Rural news	500
(7)	Crime and accidents	500
	TOTAL:	3496

TABLE III
STATISTIC OF CORPORA.

Average numb. of words in doc.	# of terms	# of symb. in term
294	106086	9

B. Feature selection

TABLE IV

DESCRIPTION OF SELECTED FEATURE MATRICES (# OF DOCUMENTS IS 3496). HERE F-SCORE IS PRESENTED AS A MEAN \pm STANDARD DEVIATION OF RESULTS BY 10 CLUSTERINGS WITH SKM ALGORITHM.

Name	# of features	F-score
LT1	2570	0.8375 \pm 0.0100
LT2	1492	0.8339 \pm 0.0059
LT3	2829	0.8326 \pm 0.0325
LT4	1399	0.8323 \pm 0.0083
LT5	1328	0.8287 \pm 0.0071
LT6	5243	0.8282 \pm 0.0448
LT7	1714	0.8266 \pm 0.0371
LT8	1740	0.8244 \pm 0.0051
LT9	2165	0.8218 \pm 0.0284
LT10	1605	0.8217 \pm 0.0080

Feature set was formed using BOW, the primary set included 106086 unique features. Four feature set reduction models (sec. II-B) were applied with an intuition that smaller number of features can describe documents as well as a complete set or even better, while decreasing duration of clustering.

150 reductions of feature set combining different values of bounds were performed:

$$\begin{aligned}\alpha_- &= \{0.001, 0.002, 0.003, \dots, 0.03\} \cdot n, \\ \alpha_+ &= \{100, 66.66, 50, 33.33\} \cdot n, \\ \alpha_{min} &= \{0.01, 0.015, 0.02, \dots, 0.2\} \cdot n.\end{aligned}$$

where n is a number of objects in corpora.

Objects with reduced number of features were clustered by KM and SKM algorithms. It was observed that SKM algorithm performs considerably better, i.e. F-score is higher from 2 to 3 times, hence experiments with reduced matrices were performed only with SKM algorithm.

Experiments show, that the best performing model is $Q_{tf}^1 \rightarrow DF \rightarrow Q_{tfidf}^2$. By eliminating part of features with this model it was found that α_+ bound does not affect number of features significantly, over 90% of primary features were eliminated by α_- bound and the best results were achieved

with $\alpha_- \in [0.011; 0.021]$. For further research, ten matrices from determined interval were selected, see table IV.

C. Clustering

All experiments were performed using matrices described in table IV, with assumption that number of clusters is known and equal to 7.

Results were compared with random baseline, where all objects are distributed at random in seven clusters. In experiments with hierarchical clustering the best results were achieved using cosine similarity and Ward linkage, while complete, single and average link perform close to random baseline (F-score).

Results of experiments with hierarchical clustering with Ward measure, KM and SKM algorithms are depicted in Fig. 1, where a) H is hierarchical algorithm with: Jaccard (J), Euclidean (E) and cosine (C) similarity measure and Ward distance between clusters (W); b) KM is K-means algorithm; c) SKM is spherical K-means algorithm; d) Random baseline is random objects distribution into 7 classes.

The results show, that algorithms almost always perform in a following sequence: SKM, HCW, HJW, HEW, KM. It clearly rejects Euclidean distance as a measure for text clustering. **Jaccard and cosine distances** perform quite well, and **SKM algorithm** achieves the best results.

IV. CONCLUSIONS

A number of experiments with different feature reduction and clustering methods were performed:

- 1) Four models for feature set reduction;
- 2) Three dissimilarity measures between objects (Euclidean, Jaccard and cosine);
- 3) Four linkages (single, average, complete and Ward linkages);
- 4) And three clustering algorithms (K-means, spherical K-means and hierarchical clustering).

All the results were evaluated using F-score index.

The best results were achieved using spherical K-means algorithm with $Q_{tf}^1 \rightarrow DF \rightarrow Q_{tfidf}^2$ feature reduction model.

Future plans include experiments with symbolic n-grams, determining number of clusters, detecting and eliminating outliers and experimenting with other clustering algorithms, e. g. density and distribution based algorithms.

In this paper cluster documents under the assumption that clusters do not overlap. However, often that is not a case. Hence, in the future we are planning to experiment with soft clustering methods.

ACKNOWLEDGEMENTS

This work was supported by Grant Nr. REP-09/2015 of the Lithuanian Research Council.

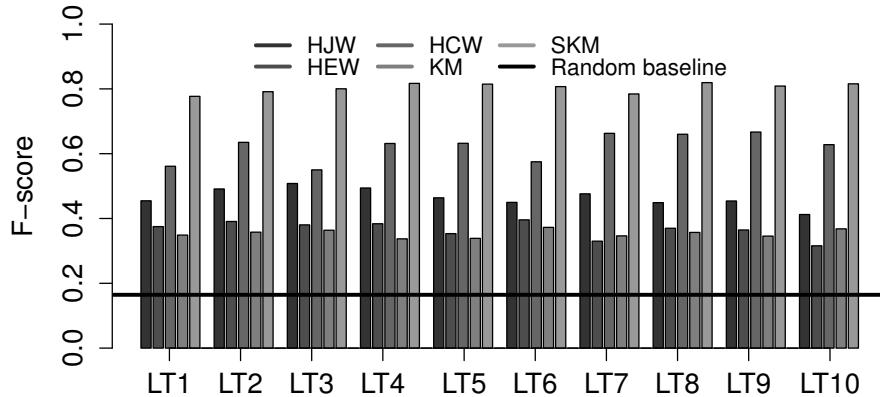


Fig. 1. F-score values of clustering results of five clustering algorithms and random baseline.

REFERENCES

- [1] C. D. Manning, P. Raghavan, and H. Schütze, *Introduction to Information Retrieval*. New York, NY, USA: Cambridge University Press, 2008.
- [2] A. K. Jain, M. N. Murty, and P. J. Flynn, “Data clustering: A review,” *ACM Comput. Surv.*, vol. 31, no. 3, pp. 264–323, Sep. 1999.
- [3] P.-N. Tan, M. Steinbach, and V. Kumar, *Introduction to Data Mining, (First Edition)*. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 2005.
- [4] M. H. Ahmed and S. Tiun, “K-means based algorithm for islamic document clustering,” 2015.
- [5] G. Ciganaitė, A. Mackutė-Varoneckienė, and T. Krilavičius, “Text documents clustering,” in *Information Technology, XIX International Conference for Master and Doctoral Students (Information Society and University Studies)*, 2014.
- [6] A. Mackutė-Varoneckienė and T. Krilavičius, “Empirical study on unsupervised feature selection for document clustering,” in *Human Language Technologies The Baltic Perspective*, 2014.
- [7] G. Ciganaitė, A. Mackutė Varoneckienė, and T. Krilavičius, “Tekstinių dokumentų klasifikavimo metodų tyrimas,” in *Information Technology 2015, XX International Conference for Master and Doctoral Students*, 2015, pp. 171–174.
- [8] J. Zobel and A. Moffat, “Exploring the similarity space,” *SIGIR Forum*, vol. 32, no. 1, Apr. 1998.
- [9] Y. Yang and J. O. Pedersen, “A comparative study on feature selection in text categorization,” in *Proceedings of the Fourteenth International Conference on Machine Learning*, ser. ICML ’97. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1997, pp. 412–420.
- [10] T. Liu, S. Liu, Z. Chen, and W.-Y. Ma, “An evaluation on feature selection for text clustering,” in *ICML*, T. Fawcett and N. Mishra, Eds. AAAI Press, 2003, pp. 488–495.
- [11] A. Strehl, E. Strehl, J. Ghosh, and R. Mooney, “Impact of similarity measures on web-page clustering,” in *In Workshop on Artificial Intelligence for Web Search (AAAI 2000)*. AAAI, 2000, pp. 58–64.
- [12] I. S. Dhillon, Y. Guan, and J. Kogan, “Iterative Clustering of High Dimensional Text Data Augmented by Local Search,” in *Proc. IEEE International Conference on Data Mining (ICDM’02)*, Dec. 2002.
- [13] M. Steinbach, G. Karypis, and V. Kumar, “A comparison of document clustering techniques,” in *In KDD Workshop on Text Mining*, 2000.

Dvimačių supjaustymo uždavinių sprendimas naudojant grafinį procesorių

Dovydas Šopa

Programų inžinerijos katedra
Kauno technologijos universitetas
Kaunas, Lietuva

Santrauka – straipsnyje analizuojama dvimačių supjaustymo uždavinių sprendimas naudojant grafinį procesorių. Pasiūlytas algoritmas naudoja evoliucinį euristinį modifikavimo tipo algoritmą, kuris naudoja euristinius algoritmus sprendiniui gauti. Pateikiamas pasiūlyto algoritmo ir realizacijos gautų rezultatų palyginimas su evoliuciniu euristiniu algoritmu, kuris skaičiavimus vykdo standartiniame procesoriuje.

Raktiniai žodžiai – daiktų išdėliojimo uždavinys, grafinis procesorius, optimizavimas, CUDA.

I. ĮVADAS

Objektų išdėliojimo uždavinys yra klasikinis optimizavimo uždavinys, kuris pirmą kartą suformuotas 1939 m. [1]. Šis uždavinys yra įdomus ne tik teoriškai, bet turi daug praktinių pritaikymų: straipsnių išdėliojimas laikraštyje taip, kad reikėtų kuo mažiau lapų detalių pjaustymas lakštose taip, kad liktų kuo mažiau nepanaudotų medžiagų; dėžių išdėliojimas sunkvežimiųose taip, kad reikėtų kuo mažiau sunkvežimių.

Yra puikiai žinoma, kad tai nedeterministiškai polinomiškai (toliau – NP) sunkus uždavinys [2]. Pirmausia šie uždaviniai pradėti nagrinėti kaip vienmačiai, o po to buvo praplesti iki daugiamatių. Klasikinė šio uždavinio formuluotė skamba taip: visos dėžutes turi ilgius 1, o daiktų ilgiai yra intervale (0; 1] kiekvienai dimensijai. Daiktus į dėžutes reikia sudėlioti taip, kad būtų panaudota kuo mažiau dėžučių. Dvimačiu atveju matematinis šio uždavinio apibarėzimas skamba taip: duotoje dvimatėje koordinacijų erdvėje A, kurios dydis R^2 , reikia surasti tokią stačiakampių aibę V, kad bet kuriai stačiakampių porai ($v; v'$) iš aibės V sankirtų $v \cap v'$ aibę būtų tuščiai [3].

NP sudėtingumo uždaviniai pasižymi tuo, kad jų sprendimui reikia didesnio nei polinominio laiko. Tai reiškia, kad šio uždavinio sprendimo trukmės apatinė riba (angl. *Lower bound*) yra $\omega(n^k)$, kur n – įvesties parametras, susijęs su duomenų kiekiu; k – bet koks skaičius.

Čia gali padėti grafiniai procesoriai (toliau – GPU). GPU gamintojai paskutiniu metu ištobulino šiuos procesorius, kad jie būtų tinkami didelės apimties skaičiavimams. Čia labai prisidėjo NVIDIA, kuri sukūrė CUDA (angl. *Compute Unified Device Architecture*). CUDA technologijos yra paremtos vienos instrukcijos daugeliui gilij pagrindu (angl. *Single instruction multiple threads*, toliau – SIMT). SIMT pasižymi

tuo, kad kelios gijos vykdo tuos pačius veiksmus su skirtingais duomenimis. Tai leidžia surūpinti laiko instrukcijoms įkrauti.

Šiame darbe nagrinėjamas objektų išdėliojimo dvimatis atvejis, kai objektai yra orientuoti stačiakampiai, negali būti vartomi, objektų sąrašas žinomas iš anksto ir lakštai yra vienodo dydžio. Bus bandoma išnaudoti CUDA suteikiamas galimybes šiam uždavinui spręsti lygiagrečiai.

Šio darbo antrame skyriuje pateiki susiję darbai. Trečiame skyriuje pateikiamas siūlomas evoliucinis euristinis algoritmas. Ketvirtame skyriuje pateikiamas atliktas tyrimas ir gauti rezultatai. Penktame skyriuje pateikiamos išvados ir tolimesni darbai.

II. SUSIJĘ DARBAI

Kadangi optimalus algoritmas NP-sunkiemis uždaviniams dar nėra sukurtas, literatūroje galima rasti daugybę straipsnių, kuriuose mokslininkai siūlo savo algoritmus. Galima pastebeti, kad geriausius rezultatus pasieka evoliuciniai euristiniai algoritmai [4, 5, 6, 7, 8]. Šių algoritmų pagrindinė idėja – priklausomai nuo situacijos pasirinkti vieną iš paprastų euristinių algoritmų.

Literatūroje galima pastebeti dvių tipų evoliucinius euristinius algoritmus: pasirinkimo [7] ir modifikavimo [8]. Kadangi modifikavimo algoritmai gali lengviau prisitaikyti prie specifinių uždavinii, šiam darbui pasirinktas modifikavimo tipo evoliucinis euristinis algoritmas.

X. Zhao ir H. Shen [9] pabandė sudaryti algoritmą, kuris objektų išdėliojimo uždavinį spręstų lygiagrečiai. Jie pasiūlė algoritmą, kuris pritaikytas 32 procesorių sistemai. Buvo sprendžiamas specifinis uždavinio atvejis (objektai kvadratiniai) ir pasiektais $\Theta(n)$ skaičiavimo laikas.

Praktikoje tokia procesorių sistema nepraktiška. Tokiems skaičiavimams būtų tikslingo naudoti CUDA. NP sudėtingumo uždavinius jau buvo bandoma spręsti naudojant GPU [10, 11]. Šie sprendimai parodo, kad CUDA puikiai tinka tokio tipo uždaviniams (rezultatai pagerėja apie 50 kartų).

Sprendimai, pritaikyti CUDA, apsiribojo genetinių algoritmų naudojimu. Kadangi nuosekliausose skaičiavimuose geriausius rezultatus pasieka evoliuciniai euristiniai algoritmai, todėl jų pritaikymas lygiagretiems skaičiavimams naudojant CUDA gali pasiekti dar geresnius rezultatus. Būtent

tai šiame darbe ir nagrinėjama – modifikavimo evoliucinis euristinis algoritmas, pritaikytas skaičiavimams CUDA.

III. SIŪLOMAS EVOLIUCINIS EURISTINIS ALGORITMAS

Kai kurie objektų išdėliojo uždaviniai pasižymi savybėmis, dėl kurių specifinės euristikos juos sprendžia greitai ir tiksliai. Tačiau nėra vienos euristikos, kuri puikiai spręstų visus galimus atvejus. Evoliucinio euristinio algoritmo idėja yra apjungti paprastus euristinius algoritmus, taip bandant išvengti kiekvieno iš algoritmų silpnų vietų.

Taip pat evoliuciniai euristiniai algoritmai problema nagrinėja iš kitos pusės. Jei euristiniai algoritmai operuoja duomenimis (objektais ir lakštais, į kuriuos šie objektai turi būti sudėti), tai evoliuciniai euristiniai algoritmai operuoja žemesnės eilės algoritmais (algoritmais, iš kurių reikia pasirinkti).

Siūlomas evoliucinis euristinis algoritmas remiasi kitų autorų mintimis [4, 5, 6, 7, 8]. Dauguma siūlomo evoliucinio euristinio algoritmo naudojamų euristinių algoritmų nėra nauji, tačiau pasirinkta specifinė algoritmų aibė, arba jie modifikuoti.

A. Naudojami euristiniai algoritmai

Vienmačiu objektų išdėliojo uždavinio atveju euristikų tikslas yra išrinkti objektą ir talpyklą, į kurią šis objektas bus dedamas. Dvimačiu atveju prisideda papildomas kintamasis – reikia rasti poziciją, į kurią bus dedamas objektas lakšte. Dėl šios priežasties naudojami dviejų tipų euristiniai algoritmai: išrinkimo (išrenka objektą ir lakštą, į kurį objektas bus dedamas) ir déjimo (išrenka poziciją lakšte, kur objektas bus dedamas). Naudojami pasirinkimo algoritmai:

- Pirmo tinkančio (angl. *first fit*) – atidarytus lakštus nagrinėja iš eilės, ieškant pirmosios, į kurį tinka objektas. Naudojamos dvi šio tipo algoritmo modifikacijos:
 - ploto mažėjimo seka – objektai nagrinėjami nuo didžiausio, iki mažiausio,
 - su iš anksto paskirtais objektais [6] – visi objektai, kurie yra didesni nei pusė lakšto dydžio, iš karto išdėliojami į lakštus.
- Kito tinkančio (angl. *next fit*) – objekto déjimui nagrinėjama tik paskutinis naudotas lakštas. Jei objektas tinka – dedama į jį, priešingu atveju – imamas naujas lakštas. Naudojamos dvi šio tipo algoritmo modifikacijos:
 - standartinis – objektai ir lakštai nagrinėjami tokia seka, kokia yra gauti,
 - su iš anksto paskirtais daiktai.
- Geriausiai tinkančio (angl. *best fit*) – lakštai išrikuojami pagal likusio laisvo ploto didėjimo seką. Šia seka jie nagrinėjami, siekiant rasti pirmą į kurį objektas telpa. Naudojama objektų ploto mažėjimo modifikacija.
- Blogiausiai tinkančio (angl. *worst fit*) – lakštai išrikuojamos pagal likusio laisvo ploto mažėjimo seką. Šia seka jie nagrinėjami, siekiant rasti pirmą į kurį telpa

objektas (lakštai nagrinėjami priešinga seką, nei geriausiai tinkančio algoritme). Naudojama objektų ploto mažėjimo modifikacija.

• Django ir Fitcho – algoritmas užpildo $1/n$ dalį lakšto naudodamas pirmo tinkančio algoritmą. Tuomet jis ieško objektų iki trijų (gali naudoti 1, 2 arba 3) kombinacijos, kuri sudarytų didžiausią plotą ir dar tilptų į lakštą. Jei kelios kombinacijos turi vienodą plotą, pirmenybė teikiama tai, kuri turi mažiau elementų. Jei ir tada yra kelios kombinacijos, pirmenybė teikiama tai, kuri turi didžiausią objektą (jei didžiausias objektas sutampa, tada ta pati salyga galioja antram arba trečiam pagal dydį objektui). Jei nei viena kombinacija negali būti įdėta į lakštą – imamas naujas. Naudojamos trys šio algoritmo modifikacijos (užpildymas iki $1/4$, $1/3$ ir $1/2$) [7]. Šiame algoritme objektai nagrinėjami ploto mažėjimo seką.

Django ir Fitcho algoritmas yra vienintelis, kuris vienu pasirinkimu gali išrinkti daugiau nei vieną objektą.

Objektų déjimui pasirinkti algoritmai, kurie orientuoti į déjimą apatiniaiame kairiajame kampe. Visi algoritmai orientuoti į panašų déjimą, kad būtų išlaikytas sprendinio stabilumas (jei dalis algoritmų détu nuo kairiojo kampo, o dalis nuo dešiniojo, tada viduryje gali likti tarpų, kurie bus nepatogios formos ir kuriuos bus sunku užpildyti). Naudojami déjimo algoritmai:

- Apatinio kairiojo kampo. Objektas įdedamas į viršutinį dešinįjį kampą. Tuomet, naudojant stūmimo žemyn ir kairėn veiksmus, objektas stumiamas tol, kol pasiekia kitą objektą arba lakšto kraštą.
- Patobulinto apatinio kairiojo kampo. Objektas, kaip ir apatinio kairiojo kampo algoritme, įdedamas į viršutinį dešinįjį kampą. Tuomet, skirtingai nuo apatinio kairiojo kampo algoritmo, objektas stumiamas ne tik į artimiausio objekto krašto. Pristūmuis objektą įki kito objekto procesas nėra nutraukiamas – tikrinama, ar už šio yra laisvos vietos, į kurią naujas objektas tilptų. Jei ši vieta randama – objektas perkeliamas per jau egzistuojantį objektą.
- Skenavimo. Šio algoritmo tikslas, skenuoti visas pozicijas iš kairės į dešinę ir iš apačios į viršų ir ieškoti pirmos, kurioje gali būti padėtas apatinis kairysis objekto kampus.

Taip iš viso gautos 27 algoritmų kombinacijos (9 pasirinkimo ir 3 déjimo algoritmai).

B. Evoliucinis euristinis algoritmas

Evoliucinį euristinį algoritmą galima aprašyti tokiais žingsniais (1 pav.):

- 1) paruošiama 20 uždavinio kopijų.
- 2) kiekvienai kopijai nustatoma dabartinė uždavinio būsena.
- 3) jei rastai būsenai jau yra pasirinkta algoritmų kombinacija – ji ir naudojama. Priešingu atveju pasirenkama atsitiktinė kombinacija iš anksčiau aprašytų.
- 4) šis procesas kartojamas, kol sudedami visi objektai.

```

while Uždavinio sprendinys nepagerinamas 10 kartų
    Paruošti 20 uždavinio kopijų
    foreach kopija įm kopijos
        while Ne visi daiktai sudėlioti
            Nustatyti dabartinę už uždavinio būseną
            if Būsena, kurių algoritmas jau pasirinktas
                Naudoti jau pasirinktą algoritmą
            else
                Atsitiktinai išsirinkti naują algoritmą
            Apskaičiuoti kiekvieno sprendinio tinkamumą
            Geriausią individų išrinkimas
            Geriausią individų rekombinacija
            Atliekama atsitiktinių individų mutacija

```

I pav. Naudojamas evoliucinis euristinis algoritmas

5) įvertinamas kiekvieno iš gautų sprendinių tinkamumas (angl. *fitness*).

6) naudojant genetinį algoritmą, išrenkama dalis geriausią sprendinį radusių algoritmų kombinaciją:

a) išrenkama 80 % algoritmų kombinaciją. Kombinacijos renkamos ruletės principu. Kiekvienam variantui, atsižvelgiant į jo tinkamumą, skiriamas intervalas intervale [0; 1]. Tuomet generuojamas atsitiktinis skaičius šiame intervale. Pasirenkama algoritmų kombinacija, kuriai priskirtame intervale yra sugeneruotas skaičius.

b) atliekama rekombinacija. Iš algoritmų kombinacijų sarašo pasirenkamos dvi atsitiktinės ir su 50 % atliekama rekombinacija. Jei rekombinacija įvyksta, tuomet atsitiktinėje vietoje sumaišomi naudoti algoritmai, jei ne – kombinacijos perduodamos tokios, kokios yra.

c) su 10 % tikimybe atliekama mutacija. Ivykus mutacijai, 20 % atsitiktinių algoritmų kombinacijų išmetama.

7) visas šis procesas kartojamas tol, kol uždavinio sprendimas nepagerinamas 10 iteracijų.

Sprendinio tinkamumas T apskaičiuojamas taip:

$$T = \frac{\sum_{i=1}^M \left(\frac{\sum_{j=1}^N S_j}{N} \right)^2}{M} \quad (1)$$

M – sunaudotų dėžučių kiekis uždaviniuui spręsti. N – išdėliotų objektų kiekis. S_j – vieno objekto plotas.

Uždavinio būsena koduojama eilute, kurią sudaro 23 simboliai:

- Pirmi 6 simboliai yra skirti likusių objektų aukščiui identifikuoti. Pagal objekto aukščio santykį su lakšto aukščiu objektai sugrupuojami į tris grupes: (0; 1/3], (1/3; 1/2] ir (1/2; 1].
- Kiti 6 (nuo 7 iki 12) simboliai yra skirti likusių objektų pločiui identifikuoti. Pagal objekto pločio santykį su lakšto pločiu objektai sugrupuojami į tris grupes: (0; 1/3], (1/3; 1/2] ir (1/2; 1].
- Kiti 8 (nuo 13 iki 20) simboliai yra skirti likusių objektų plotui identifikuoti. Pagal objekto ploto santykį su lakšto plotu objektai sugrupuojami į 4 grupes: (0; 1/4], (1/4; 1/3], (1/3; 1/2] ir (1/2; 1].

- Paskutiniai 3 (nuo 21 iki 23) simboliai yra skirti likusių nejdėtų objektų kiekiu santiui su pradiniu objektų kieku identifikuoti.

Kiekvienas iš šių identifikatorių gauna reikšmę pagal jo atitikimą vienam iš intervalų procentais. Pirmoms trims kategorijoms šifruoti naudojamas 2 simbolių kodavimas, kur „00“ atitinka intervalą [0; 10], „01“ – (10; 25], „10“ – (25; 50], „11“ – (50; 100]. Likusių objektų šifravimui naudojamas 3 simbolių kodavimas, kur „000“ atitinka intervalą [0; 12,5], „001“ – (12,5; 25], „010“ – (25; 37,5], „011“ – (37,5; 50], „100“ – (50; 62,5], „101“ – (62,5; 75], „110“ – (75; 87,5], „111“ – (87,5; 100].

C. Uždavinimo lygiagretinimas

Kadangi didžiajų sprendimo laiko dalį sudaro tinkamos pozicijos lankste ieškojimas, todėl nuspresta būtent šią dalį realizuoti naudojant CUDA technologijas. Tam visi 3 objektų déjimo algoritmai pritaikyti lygiagretiemis skaičiavimams.

Apatinio kairiojo kampo ir patobulintą apatinio kairiojo kampo algoritmus galima lygiagrečiai vykdyti tokiamis gijų skaičiuje, kiek objektų tuo metu yra lankste. Stūmimo žemyn žingsnyje kiekviena gija pirmiausiai identifikuojama, ar objeketas aktualus paieškai (bent dalis objekto turi būti po dedamu objektu). Tuomet, kiekviena gija pažymi jos tikrinamo objekto užimamą aukščio intervalą (šiame intervale jau yra objeketas, todėl naujas nebegali būti dedamas). Galiausiai, viena gija (esant dideliems objekto ir lanksto matmenų skirtumams, šis veiksmas tai pat padalinamas keliomis gijos po lygiai) peržiūri pažymėtą informaciją ir patikrina, ar yra nepažymėtas reikiama dydžio intervalas. Atitinkami veiksmai vykdomi ir objektą stumiant į kairę.

Skenavimo algoritmo įgyvendinimas šiek tiek skiriasi. Čia gijų kiekis pasirinktas pagal lanksto aukštį. Kiekviena gija nagrinėja galimybę, kad apatinius kairiuoją objekto déjimo kampu bus viena iš jos nagrinėjamos eilutės pozicijų. Tinkamo taško eilutėje radimui taikomas tokis pats principas, kaip ir vietai, į kurią objektas dedamas apatinio kairiojo kampo ir patobulinto apatinio kairiojo kampo algoritmuose, nustatyti.

IV. REZULTATAI

Algoritmas realizuotas naudojant .NET 4.6 technologijas bei *Cudafy.NET* biblioteką .NET kodui transformuoti į CUDA kodą. Tyrimai daryti nešiojamame kompiuteryje, kuris turi Intel® Core™ i7-4710HQ 2,5 GHz procesorių, 8 GB darbinės atminties (RAM) ir NVIDIA GeForce GTX 850M GPU. Palyginimui pasirinktas C. Bluma ir V. Schmido pasiūlytas algoritmas [8]. Šis algoritmas šiuo metu pasiekia vienus iš geriausiu rezultatų. Pasirinkti pirmu penkių klasių (I–V) duomenys. Informacija apie šias duomenų klasses pateikta I lentelėje. Kiekvienoje klasėje objektų kiekis yra iš aibės {20, 40, 60, 80, 100}. Kiekvienai klasei ir kiekvienam objektų kiekiui pateikiama po 10 duomenų rinkinių. Taip iš viso gaunama 250 duomenų rinkinių algoritmams palyginti.

Žinoma, negalima tiesiogiai palyginti C. Bluma ir V. Schmido gautų sprendimo laikų su šiame darbe pasiūlyto algoritmo sprendimo laikais, kadangi kompiuterių parametrai

ženkliai skyrėsi. Šie laikai pasirinkti kaip atskaitos taškas, kurj CUDA technologijomis realizuotas algoritmas turėtų įveikti.

I LENTELĖ. DUOMENŲ RINKINIŲ KLASĖS

Klasė	Objekto plotis	Objekto aukštis	Lakšto plotis	Lakšto aukštis
I	[1; 10]	[1; 10]	10	10
II	[1; 10]	[1; 10]	30	30
III	[1; 35]	[1; 35]	40	40
IV	[1; 35]	[1; 35]	100	100
V	[1; 100]	[1; 100]	100	100

Gauti rezultatai apibendrinti II lentelėje. Kiekvienoje eilutėje pateikiama 10 duomenų rinkinių jungtiniai rezultatai. Pirmame stulpelyje pateikiamas objektų kiekis klasėje, antrame – geriausias iki šiol žinomas šių 10 duomenų rinkinių sprendimo rezultatas (LB, angl. *Lower bound*). Rezultatų palyginimas pateiktas dviem stulpeliais: pirmajame pateikiamas algoritmo gautas rezultatas, o antrajame – laikas, per kurį rezultatas pasiekta. Paskutinėje eilutėje pateikiama visų reikalingų lakštų kiekijų suma ir vidutis vieno duomenų rinkinio vykdymo laikas.

Rezultatų patikimumui užtikrinti buvo imtasi kelių priemonių. Visų pirma, naudojamas kompiuteris buvo atjungtas nuo interneto. Taip pat buvo išjungtos visos programos ir procesai, kurių nereikia normaliam operacinės sistemos darbui užtikrinti. Galiausiai, skaiciavimai su kiekvienu duomenų rinkiniu buvo vykdomi bent po 2 kartus. Jei abiejų vykdymų laikai buvo panašūs – imamas šių laikų vidurkis. Priešingu atveju skaičiavimai su duomenų rinkiniu buvo vykdomi trečią kartą. Tuomet buvo atsisakoma labiausiai nukrypusio rezultato ir pateikiamas 2 artimesnių rezultatų vidurkis.

Palyginus rezultatus galima pastebeti, kad CUDA vykdyti skaiciavimai nusileidžia C. Bluma ir V. Schmido gautiems rezultatams. Pirmiausia tai privertė suabejoti realizacijos korektiškumu. Dėl to buvo atlirkas papildomas realizuotos programos tyrimas naudojant ANTS Performance Profiler. Šis įrankis leidžia išanalizuoti, kurios programos vietos vykdomos daugiausiai laiko. Atlirkus šią analizę pastebėta, kad apie 60 % laiko programa praleidžia siūsdama duomenis tarp CPU ir GPU. Tai visiškai neefektyvu. Jei atmetumėme šiuos 60 % laiko, tada vidutinis vykdymo laikas nuo 5,39 s nukristų iki 2,16 s. Šie rezultatai verčia susimąstyti apie reikalingus realizacijos pakeitimus.

V. IŠVADOS IR TOLIMESNI DARBAI

Ši evoliucinio euristinio algoritmo realizacija yra neefektyvi. Dėl nuolatinio duomenų perdavimo tarp CPU ir GPU labai smarkiai nukentėja vykdymo laikas. Reikia sumažinti bendravimą tarp CPU ir GPU. Tam reikia ne tik atskirų programos dalii (dabariniu atveju tik déjimo algoritmu), bet viso evoliucinio euristinio algoritmo pritaikymo skaiciavimas GPU. Tai ir bus tolimesnis darbas.

Taip pat reikėtų atsisakyti *Cudafy.NET* bibliotekos ir tiesiogiai naudoti CUDA teikiamą sąsają. Sistemos kūrimo

metu pastebėtos kelios bibliotekos klaidos, dėl kurių sugeneruotas kodas kartais būna neteisingas arba neoptimalus.

II LENTELĖ. REZULTATU PALYGINIMAS

Uždavinys	LB	EA-LGF [8]		Šiame darbe pasiūlytas	
		Rezultatas	Laikas (s)	Rezultatas	Laikas (s)
I klasė					
20	71	71	0,00	71	0,00
40	134	134	0,00	134	0,00
60	197	200	0,01	200	0,00
80	274	275	0,00	275	0,00
100	317	317	0,00	317	0,00
II klasė					
20	10	10	0,00	10	0,01
40	19	19	0,00	19	0,01
60	25	25	0,00	25	0,01
80	31	31	0,00	31	0,01
100	39	39	0,00	39	0,01
III klasė					
20	51	51	0,02	51	0,04
40	92	94	0,01	94	0,05
60	136	139	0,27	139	1,02
80	187	189	20,68	189	39,85
100	221	224	26,17	223	44,21
IV klasė					
20	10	10	0,00	10	0,02
40	19	19	0,00	19	0,02
60	23	23	12,18	25	0,03
80	30	31	0,00	32	14,36
100	37	37	0,00	38	0,04
V klasė					
20	65	65	0,00	65	0,01
40	119	119	0,03	119	0,04
60	179	180	0,14	180	0,17
80	241	247	0,03	247	0,03
100	279	284	27,33	284	34,62
Rezultatai	2806	2833	3,47	2836	5,39

LITERATŪRA

- [1] L. V. Kantorovich, „Mathematical Methods of Organizing and Planning Production“, „Management Science“, Maryland, Vol. LX, No. 4, INFORMS, p. 366–422, 1960.
- [2] M. R. Garey ir D. S. Johnson, „Computers and Intractability: A Guide to the Theory of NP Completeness“, New York: W.H. Freeman and Company, 1979.

- [3] A. Quilliot ir H. Toussaint, „About Casting 2D Bin Packing into Network Flow Theory“, HAL, 2010.
- [4] H. Terashima-Marín, C. J. Farias-Zárate, P. M. Ross ir M. Valenzuela-Rendón, „Comparing Two Models to Generate Hyper-heuristics for the 2D Regular Bin Packing Problem“, įtraukta į „Proceedings of the 9th annual conference on Genetic and evolutionary computation“, New York, ACM, 2007, p. 2182–2189.
- [5] S. Hong, D. Zhang, H. C. Lau, X. Zeng ir Y. Si, „A hybrid heuristic algorithm for the 2D variable sized bin packing problem“, „European Journal of Operational Research“, Vol. CCXXXVIII, No. 1, p. 95–103, 2014.
- [6] M. Quiroz-Castellanos, L. Cruz-Reyes, J. Torres-Jimenez, C. Gómez S., H. J. Fraire Huacuja ir A. C. Alvim, „A grouping genetic algorithm with controlled gene transmission“, „Computers & Operations Research“, Vol. LV, p. 52–64, 2015.
- [7] E. López-Camacho, H. Terashima-Marín, P. Ross ir G. Ochoa, „A unified hyper heuristic framework for solving bin packing problems“, „Expert Systems with Applications“, Vol. XLI, No. 15, p. 6876–6889, 2014.
- [8] C. Blum ir V. Schmid, „Solving the 2D bin packing problem by means of a hybrid evolutionary algorithm“, „Procedia Computer Science“, Vol. XVIII, p. 899–908, 2013.
- [9] X. Zhao ir H. Shen, „A Parallel Algorithm for 2D Square Packing“, įtraukta į „Proceedings of the 2013 International Conference on Parallel and Distributed Computing, Applications and Technologies“, Washington, IEEE Computer Society, 2013, p. 179–183.
- [10] V. Boyer, D. El Baz ir M. Elkhiel, „Solving knapsack problems on GPU“, „Computers & Operations Research“, Vol. XIX, p. 42–47, 2012.
- [11] M. Calid Feier, C. Lemnaru ir R. Potolea, (2010), „Solving NP-Complete Problems on the CUDA Architecture using Genetic Algorithms“, „10th International Symposium on Parallel and Distributed Computing“

Research and implementation of task and project management system based on Ext JS framework

Šarūnas Slanyš
Kaunas University of Technology
Kaunas, Lithuania
sarunas.slanysh@gmail.com

Šarūnas Packevičius
Kaunas University of Technology
Kaunas, Lithuania
sarunas@ieee.org

Abstract — Task and project management is an important task of each company. Such companies need task and project management software to efficiently manage processes, focused on the successful, rapid and high quality implementation of the project. This paper researches tasks and projects management system software design and development process, in order to give the company an innovative, quick and user-configurable system. The system provides functions such as management of company's employees, task scheduling, employee employment tracking, task calendar, document management, etc. Task and project management system can reduce enterprise project management costs, improve project quality and increase enterprise productivity. Based on the Ext JS framework and MVC architecture, implemented system provides responsive user interface and feature-rich cross-platform web application.

Keywords — Task management; Project management; MVC; Ext JS; REST;

I. INTRODUCTION

At present, companies offering employees a flexible work schedule or the ability to work in a comfortable and always changing workplace, encounter problems, how to manage employees employment and maintain the company's productivity [1]. Successful and comfortable work planning and distribution of work is an important task of each company. Such companies need task and project management software to effectively manage processes, focusing on the unique achievement of the objective (e.g. the creation of a product or service) that has a beginning and an ending, and limited resources [2]. All business processes must be managed at the lowest possible cost to manage projects, tasks, documents, and other daily events taking place in the company.

This paper analyses the task and project management system development, using MVC (Model-View-Controller) architecture. This architecture allows separating presentation, content and logic layers to adapt to the complex user interface design requirements for the development of an online web system [3]. REST service coupled with the MVC architecture simplifies the query communication between the user and the server-side part, increases query and programming code readability and legibility of the modules, simplifies the system by providing simpler system scalability.

Ext JS framework provides comprehensive and convenient data management functionality that implements asynchronous communication between the client requests and

server responses reduces the queries response and data display time [4]. Sending requests in JSON format, allows saving transferred data traffic.

II. SYSTEM REQUIREMENTS ANALYSIS AND FEATURES

A. System requirements analysis

Along with the fast pace of work and urgent task changes the system must be integrated into other systems used by users on various convenient ways. Sending emails to user mailboxes and messages to consumers' mobile devices reduces the time from the point of task submission to the moment when task actually reaches a consumer.

Task and project management system has two kinds of user roles: project managers and tasks executors. Project managers have the highest rights and have the ability to manage the company's database of contacts, projects, employees, also create tasks to other users, manage tasks files, export / import tasks in XLSX, CSV format. Also, project manager's role inherits all rights of task executor. Task executor role dedicated to the assigned tasks preview, tasks execution or submitting additional information about the task. Tasks are divided into five stages: design, execution, testing, delivered to the client and completed. Notifications about task creation or changes of the task status are sent by e-mail with all required information about the task.

B. System features

According to given requirements for the system there is a need for internet based application designed for task and project management through a web browser. To create system there is necessary to implement these packages: user interface module, a project manager interface module, email sending module and communication with the database module. Users can access system only after successful authorization with the personal data in order to store all users performed actions in the database. As long as users have to use the system on a daily basis, there is implemented an intuitive user interface that allows users to arrange objects according to their needs to reach system comfortability and usability. System features design is showed in Figure 1.

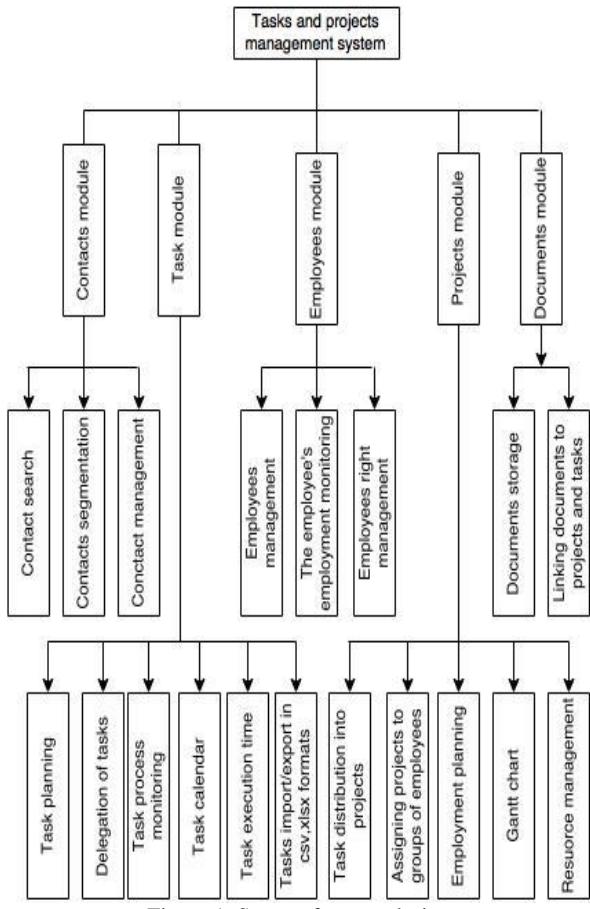


Figure 1. System features design

According to Figure 1 system can be distinguished into five user interface modules: contacts module, tasks module, employees module, projects module, documents module.

III. TECHNOLOGY SELECTION

A. MVC architecture

MVC operation is based on the separation of concerns (SoC) principle [5]. This means that the program is decomposed into the different parts that their functionality or requirements do not overlap. This principle is applied in many places. One example would be HTML and CSS. HTML is designed to specify the document structure while CSS defines how the document should look like. In object-oriented programming languages this principle is reflected by splitting tasks into objects and their methods.

The principle of the MVC gets the separation of logic, presentation and content layers. The three parts of MVC are responsible for their own activity:

- *Model* – data management or application unit of certain area.

- *View* – gives the appearance of a program model suitable for the user to understand (forms, lists, etc.). Web application is nothing more than the generated HTML / CSS code.
- *Controller* – responsible for the model management and view changes according to appropriate user interaction like button press, text entry.

B. ExtJS framework

Ext JS framework is an open source JavaScript library that provides extensive component library which makes it possible to use asynchronous Ajax requests [6], DOM, CSS, HTML and other technologies for building feature-rich cross-platform Web applications targeting the desktop, tablets, and smartphones. Ext JS leverages HTML5 features on modern browsers while maintaining compatibility and functionality for legacy browsers.

Ext JS provides hundreds of high-quality UI widgets that are meticulously designed to meet both the needs of the easiest web applications, as well as the most complex web applications. Ext JS templates and layout manager gives complete control of the screen, regardless of the devices and screen sizes. Advanced charting package allows visualize large amounts of data. This system includes a robust data packet that can use data from any backend data source. Ext JS also offers a number of out-of-the-box themes and full theming support. Ext JS also includes the availability of a package (ARIA) for people with disabilities to help Section 508 compliance [7].

There are many different JavaScript frameworks available. JavaScript MVC architecture frameworks comparisons, according to important features which are needed for task and project management system development are shown in table 1.

TABLE I. JAVASCRIPT FRAMEWORKS COMPARISON

	AngularJs	ExtJs	Backbone	React	Ember
UI binding	✓	✓		✓	✓
Reusable components	✓	✓	✓	✓	✓
Routing	✓	✓	✓		✓
Composed views		✓			✓
Grid	✓	✓			
Rich text editor		✓		✓	
Charts		✓			
Cross browser support	✓	✓	✓	✓	✓
Offline storage	✓	✓			
Developer tools		✓			

Ext JS framework chosen to establish a system adapted for convenient everyday use, depending on the usability of the system. Ext JS framework provides many objects that can be modified according to each user's needs. Integrating some of these objects to task and management system we obtain an intuitive, user-configurable user interface.

C. REST server

Representational State Transfer (REST) is architectural style for distributed hypermedia systems. REST ignores the details of component implementation and protocol syntax in order to focus on the roles of components, the constraints upon their interaction with other components, and their interpretation of significant data elements [8, 9].

This paper uses simplified open source REST service designed for PHP programming language, which was written by Jacob Wright [10].

REST is a way of formatting the URL of tasks and projects management web application. For example, if application is sending HTTP GET request to server with URL *domain.com/tasks* – server will return JSON object containing data about all tasks. Also if we are sending HTTP GET request with *domain.com/tasks/:id* – server returns only one task by given ID. In the same way we can use HTTP POST requests to create new records, HTTP PUT requests to update record by given ID and HTTP DELETE requests to delete record by given ID.

D. Other used technologies

Other technologies and the programming language that was used during the development of the task and project management system are listed below:

- *Programing languages* – PHP 5.5.10, JavaScript
- *Web server* – Internet Information Services (IIS) 7
- *Database* – Microsoft SQL server 2008
- *Operating system* – Microsoft Windows Server 2008 Service Pack 2

IV. DEVELOPED SYSTEM

Task and project management system goal – improve enterprise tasks management process. At any time, while using system we can see currently running processes in the company, state of running tasks and their progress. Also the employment of employees can be evenly distributed.

A. System design

The system consists of the IIS server, SQL Server database and mail server. In the IIS server there is placed all software code.

A key element of the system is placed in the IIS server - the user interface module based on Ext JS framework and using the MVC architecture combines the working environment of two user roles. Using that kind of the architecture of the system can always be extended with additional modules. Also, in the IIS server hosted Ext JS framework source files and programmed REST API, which includes database management methods and XLSX / CSV file

generation, parsing methods. Connections between system object is showed in Figure 2.

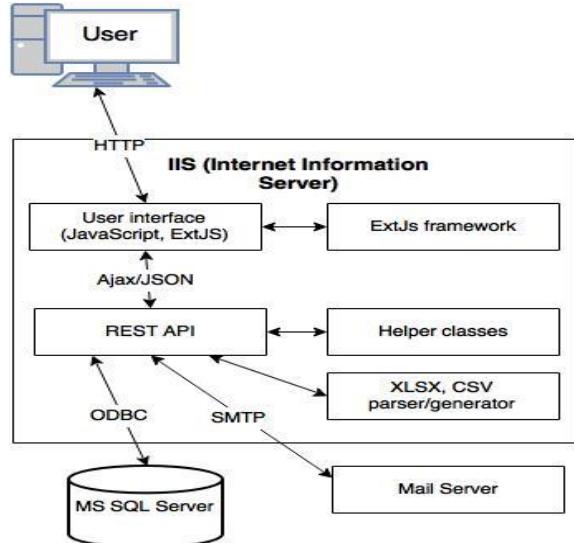


Figure 2. Connections between system objects

User using computer browser connects to the system via HTTP protocol. The user interface communicates with the server using Ajax requests, which enables to make queries asynchronously. The data transmitted in the JSON format. The server provides query the database using ODBC API, emails sent via the SMTP protocol.

B. Task lifecycle

Tasks are the main object of tasks and projects management system. Task lifecycle and state diagram is showed in Figure 3.

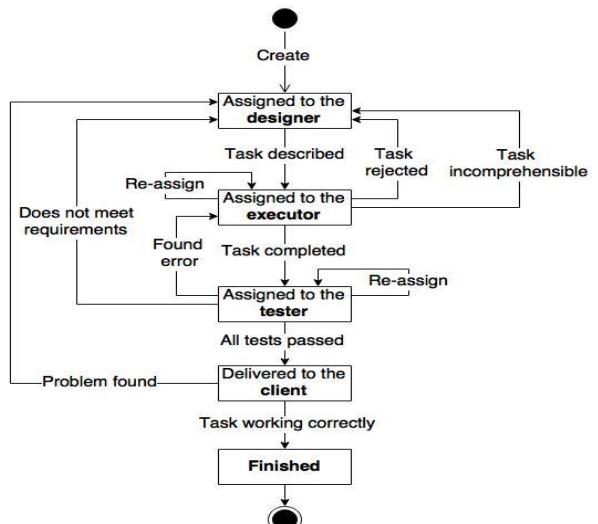


Figure 3. Task lifecycle diagram

- In this paper task lifecycle is divided into five stages:
- *Assigned to the designer* – status in which the designer designs the task to extract and provide all the necessary information about a task that will be required during the execution of the task. The designer also assigns a person responsible for the task execution and testing. At the end of the design, the designer transmits task to executor.
 - *Assigned to the executor* – state in which the task is executed. If executor cannot perform the task or task is incomprehensible according to the description then it can be returned to design status. Also, the task can be re-assigned to another employee who can complete the task. If the task has been completed successfully, then task can be transferred to the testing stage.
 - *Assigned to the tester* – status in which task testing is performed. If error is found during testing, the task is returned to the executor. If there is observed that the task does not meet requirements or specifications, the task is returned to the design state. After successful testing process, the task is being delivered to customer.
 - *Delivered to the client* – state when the task has been completed, tested and given to the client for further use. If customer notices inaccuracies, the task is returned to design stage with customer comments and all the task process starts from the beginning. If the customer does not provide observations about the incorrect functioning or customer confirms the correctness of the task operation – the task is completed.
 - *Finished* – state when task is completed and marked as finished. Finished tasks are stored in the system database, in order to display project and task history and statistics related to task.

V. CONCLUSIONS

- Developed task and project management system allows companies manage and monitor the company's task and project management process, improving the traceability of tasks and increasing the company's productivity.
- According to requirements developed an online system, accessible through a computer, tablet or smartphone web browser with maintaining compatibility and functionality for legacy browsers.
- The system is realized based on Ext JS framework using the MVC architecture, which realizes the feature-rich cross-platform web application. Feature-rich widgets provide users comfortable responsive workplace.
- In this paper, the task lifecycle is divided into five main stages: design, execution, testing, delivered to

client and finished. Every task stage have assigned responsible person, who is responsible for particular stage quality.

- A system can be used in various companies for tasks employment planning, tasks process monitoring, performed work history trail. Email services inform system users about new tasks and task status changes, to ensure faster task information delivery.

REFERENCES

- [1] Gamze Karayaz, Charles B. Keating, Morgan Henrie, „Designing Project Management Systems“. System Sciences(HICSS), 2011, pp. 1-10.
- [2] Rahat Iqbal, Richard Rider, Nazaraf Shah, Anne James, „Project management system review and redesign using user-centred design methodology“. Computer Supported Cooperative Work in Design (CSCWD), 2011.
- [3] Hamid Mccheick, Yan Qi, „Dependency of components in MVC distributed architecture“. Electrical and Computer Engineering (CCECE), 2011, pp. 139-145.
- [4] Libin Yang, Yilai Zhang, Hailin Li, „Information Search Based on Test Mining and EXTJS“. Software Engineering, 2009, pp. 386-391.
- [5] Liu Xudong, Xu Xiaofei, Zhan Dechen, He Ting, „A Novel Approach for Developing Web-Based Enterprise Information Systems“. Computer Science and Software Engineering, 2008, pp. 548-551.
- [6] Sun Zhaoyun, Zhang Xiaobo, Zhao Li, „The Web asynchronous communication mechanism research based on Ajax“. Education Technology and Computer (ICETC), 2010, pp. 370-372.
- [7] Rosalind Rogoff, „Making electronic information accessible to everyone“. Professional Communication Conference, 2001, pp. 231-236.
- [8] Roy Thomas Fielding, Richard Taylor, „Principled design of the modern Web architecture“. Software Engineering, 2000, pp. 407-416.
- [9] Roy Thomas Fielding, Architectural Styles and the Design of Network-based Software Architectures. Doctoral dissertation, University of California, Irvine, 2000.
- [10] Rest server. Internet access: <https://github.com/jacwright/RestServer>

ASAS – lietuvių kalbos anaforų sprendimo analizės ir išvertinimo prototipas

Arūnas Čiuksys

Kauno technologijos universitetas
Informacijos sistemų katedra
Kaunas, Lietuva
arunas.ciuksys@ktu.edu

Rita Butkienė

Kauno technologijos universitetas
Informacijos sistemų katedra
Kaunas, Lietuva
rita.butkiene@ktu.lt

Santrauka – kasdien augant skaitmeninės informacijos kiekiui, vis aktualesnė tampa semantinė paieška. Semantinė paieška padeda iš nestruktūruotų tekstu išgauti daugiau informacijos, nei atliekant paiešką pagal raktinius žodžius. Vienas iš kokybiškesnė semantinė paieška leidžiančių išgavendinti komponentų yra anaforų sprendimas. Anaforų sprendimo sistemos lietuvių kalbai šiuo metu yra kuriamos. Tačiau įrankių, leidžiančių efektyviai išvertinti sukurta ar kuriamas anaforų sprendimus, nėra. Todėl buvo nuspresta sukurti kompiuterizuotą įrankio prototipą ASAS, kuris ir pristatytas šiame straipsnyje. ASAS suteikia galimybę išvertinti anaforų sprendimus, sukuriant anaforų anotavimo tekstyną ir jo etaloną, automatiškai palyginant anaforų sprendimo sistemos sužymėtas anaforų anotacijas tekstyne su etalonu.

Raktiniai žodžiai – ASAS, anaforos, lietuvių kalbos anaforų sprendimo analizė ir išvertinimas, semantika, natūralios kalbos apdorojimas, skaitmeninė informacija, koreferencijos.

I. ĮVADAS

Nuolat augant internete publikuojamos informacijos kiekiui, prasminka paieška nestruktūruotose tekstuose tampa vis svarbesnė. Prasminges, ne vien tik raktažodžiais grįstos paieškos išgavendinime svarbų vaidmenį atlieka natūralios kalbos apdorojimo (NKA) sistemos, t. y. sistemos, kurios automatiškai išanalizuja tekstą, nustato jo struktūrą, žodžių morfolingines savybes, junginius, sakinių sintaksę, įvardytas esybes ir kitas teksto ir jo dalių savybes. Turint patikimai veikiančias NKA sistemas galima semantinės paieškos tekstuose plėtoti sprendimus, leidžiančius tiksliau atsakyti į vartotojų pateiktas užklausas, sumažinti rankinio paieškos darbo apimtis.

Informacijos apie NKA sprendimus vyraujančioms pasaulio kalboms galima rasti jau nuo 1970 m. Tuo tarpu NKA sprendimai lietuvių kalbai pradėti kurti palyginti nesenai ir ryškesnė pažanga stebima per pastaruosius keletą metų. 2015 metais VDU ir KTU vykdyto projekto Semantika-LT („Lietuvių kalbos sintaksinės – semantinės analizės sistema tekstynei, lietuviškam internetui ir viešojo sektorius taikymams“, vykdytas pagal Ekonomikos augimo veiksmų programos 3 prioriteto „Informacinių visuomenės visiems“ išgavendinimo priemonę Nr. VP2-3.1-IVPK-12-K „Lietuvių kalba informacinėje visuomenėje“) metu buvo sukurtu lietuvių kalbos sintaksinės semantinės analizės informacinių sistemas (LKSSAIS), teikianti vartotojams NKA ir semantinės paieškos tekstuose paslaugas. LKSSAIS semantinė paieška grindžiama tekstu morfolingine, sintaksine ir įvardytų esybių informacija, kurią pateikia atitinkami LKSSAIS NKA komponentai. Tačiau

ši informacija leidžia realizuoti tik ribotą informacijos atpažinimą. Geresnių rezultatų būtų galima tikėtis, jei semantinė tekstu analizei būtų pateikta informacija apie anaforas. Anaforų sprendimas yra svarbi NKA sistemos dalis.

Anafora yra vienas iš koreferencijos ryšio tipų, kurį ji detalizuoja: išskiriamas anaforos pirmtakas ir objektas. Anaforų sprendimo [1] metu atpažįstami skirtinėse teksto vietose esantys to paties subjekto paminėjimai, kurie susiejami anaforos ryšiu. Anaforos objektai, kaip atskiri žodžiai, dažnai neturi jokios prasmės (pvz., įvardinės anaforos objektai), bet susiejus juos su anaforos pirmtaku, jie igauna prasmę, kurią suteikia anaforos pirmtakas. Dėl to, apdorojant tekštą, yra galimybė iš jo išgauti daugiau informacijos. Pavyzdžiu: „Tomas šiandien nebuvu mokykloj. Jis sirgo.“ Žodžiai „Tomas“ ir „Jis“ sudaro anaforą (siejasi anaforos ryšiu). Be anaforų sprendimo negalėtume žinoti, kodėl šiandien Tomas nebuvu mokykloj arba, kas sirgo. Be natūralios kalbos apdorojimo netenkama daug panašios informacijos.

Anaforų sprendimo sistemos yra labai sudėtingos [2]. Pirmasis bandymas sukurti anaforų sprendimą lietuvių kalbai Semantika-LT projekte parodė, kad tokio sprendimo kūrimas turi būti laipsniškas, nuolat tikrinant ir vertinant sukurto algoritmo kokybę. Norint atlirkti tai efektyviai, reikia turėti metodiką ir analizę bei vertinimą automatizuojančius įrankius.

Antrame skyriuje aprašomi kriterijai, kuriais turėtų pasižymeti anaforų sprendimams vertinti naudojami įrankiai, ir jų pagrindimas. Trečiame skyriuje pristatomas sukurtais įrankio prototipas ASAS.

II. MOTYVACIJA

Kuriant lietuvių kalbos anaforų sprendimo sistemas svarbu sudaryti sąlygas išvertinti, ar priimti sprendimai dėl anaforų atpažinimo algoritmulį duoda gerus rezultatus. Efektyvesniams vertinimui praverstę kompiuterizuotas įrankis, kuris įgalintų sudaryti anaforų anotavimo etaloninį tekstyną ir jį palyginti su anaforų sprendimo sistemos rezultatais, bei išvestų išvertinimo matus.

A. Sprendimo išvertinimas, etalonų sudarymas ir bendra prieiga

XX a. 9-ojo dešimtmečio pabaigoje buvo organizuojamos anglų kalbos teksto automatinio natūralios kalbos apdorojimo sistemų konferencijos MUC (angl. *Message Understanding Conference*). Pagrindinė konferencijų paskirtis buvo išvertinti ir

tarpusavyje palyginti teksto apdorojimo sistemas, joms atliekant tas pačias užduotis, t. y. apdorojant tuos pačius tekstus siekiant gauti tuos pačius rezultatus [3], [4]. Sprendimams vertinti buvo pasirinkta naudoti dydžius R (išsamumas), P (tikslumas) ir F-vertė.

Išsamumas R (1), kilęs iš signalų teorijos, parodo santykį tarp gautų teisingų rezultatų C ir visų tekste esančių rezultatų T kiekui.

$$R = C / T \quad (1)$$

Tikslumas P (2) parodo santykį tarp tekste gautų teisingų rezultatų C ir visų gautų rezultatų F kiekui.

$$R = C / F \quad (2)$$

Šioms charakteristikoms papildomai vertinti naudojama F-vertė (angl. *F-measure*) (3). Tai yra harmoninė išraiška tarp tikslumo (P) ir išsamumo (R).

$$F = (2 \times P \times R) / (P + R) \quad (3)$$

Automatinį sprendimų įvertinimas apdorojant skirtingą tekstą negalėjo būti lyginamas, nes skirtingi algoritmai vertinant skirtingo ilgio ar žanro tekstus gali kiekvieną kartą pasirodyti skirtingai. Todėl *MUC* konferencijose automatiniams sprendimams įvertinti reikėjo pateikti tas pačias užduotis. Tuo tikslu buvo sukurtas tekstynas ir sudarytas tekstyno etalonas (angl. *gold standard corpus*). Norint išvengti netikslumų dėl minėtų problemų, tekstas turėjo būti ilgas ir turintis daug skirtingų žodžių. Trečiosios *MUC* konferencijos metu buvo naudojamas tekstu rinkinys, kurį sudarė 400 000 žodžių, iš kurių 18 000 buvo skirtingi.

Konferencijų metu pasirinkti įverčiai ir sudaryti tekstyne, iš dalies ar pilna apimti naudojami iki šiol. Jais įvertinamos įvairios anglų ir kitų kalbų automatinės teksto apdorojimo sistemos. Naudojami ne tik minėtieji įverčiai, bet ir nuorodų dažnumas RR (4) – panašus į tikslumą matas, bet vietoje vardiklyje naudojamo sistemos gautų teisingų rezultatų kiekio imama visų tekste esančių rezultatų (T) ir sistemos blogų rezultatų (E) kiekijų suma.

$$RR = C / (T + E) \quad (4)$$

Šiuo metu anglų kalboje vertinant įvairius sprendimus yra naudojami *MUC* konferencijų tekstu rinkiniai [5], [6]. Tokiu būdu sprendimai gali būti palyginami pavieniui (angl. *off-site*).

Dėl skirtumų tarp kalbų anaforų sprendimo sistemos negali būti daugiakalbės ir lietuvių kalbai anglų kalbos tekstyne negali būti naudojami. Bet lietuvių kalbos tekstyne, kuriam sudarytas anaforų anotavimo etalonas, šiuo metu nėra.

Tekstyne ir jo etalonu sudarymo procesas yra labai sudėtingas procesas: surinkti tinkamo ilgio ir sudėtingumo tekstu, turinčius daug skirtingų žodžių, sudaryti jų anotavimo etalonus, sužymint anaforas tekste rankiniu būdu, gali tik lietuvių kalbos specialistas, o norint sprendimus palyginti tarpusavyje skirtingi kūrėjai turi naudoti tą patį tekstyną. Todėl kompiuterizuotą įrankį tikslinę naudoti ne tik tekstynei ir jo etalonui sudaryti, bet ir galimybei skirtingiemis anaforų sprendimo sistemų kūrėjams prieiti prie sudaryto tekstyno. Priėjimas prie vieningos sudarytos duomenų bazės įgalina anaforų sprendimo įvertinimą pavieniui. Be to, sutaupo resursų,

nes nereikia sudarinti tekstyno ir jo etalonu. Antra, net ir naudojant bendrai prieinamus tuos pačius tekstyndus, norint gauti sprendimų įvertinimo dydžius tikslinę naudoti kompiuterizuotą įrankį – verčių apskaičiavimas rankiniu būdu būtų ypač imlus darbui ir laikui.

B. Skirtingi anaforų tipai

Anaforos yra skirtomos į tipus. Pagal anaforos objekto tipą ir dėl skirtumų tarp kalbų, jos kiekvienoje kalboje klasifikuojamos skirtingai. 2014 m. buvo pasiūlyta pirmoji lietuvių kalbos anaforų taksonomija [7], kurioje buvo išskirti 3 anaforų tipai: morfologinio, leksinio semantinio ir dalykinės srities tipo. Tipai buvo suskirstyti į daugiau potipių. Pavyzdžiu, tikrinio daiktavardžio – *Tomas, Petras*; vietos prieveiksmio – *čia, ten*; asmeninio įvardžio – *aš, tu, jie, jos*. Daugelis kitoms kalboms esamų anaforų sprendimų nesugeba atpažinti visų anaforų tipų.

Pirmieji anaforų sprendimai, kuriami lietuvių kalbai, yra skirtingi įvardinėms anaforoms atpažinti. Svarbu, kad sprendimo įvertinimo įrankis suteiktų galimybę, kuriant etaloninį tekstyną, išskirti anaforų tipus ir įvertinti algoritmą sprendžiant tik tam tikro tipo anaforas. Be to, kol nėra nusistovėjusios ir bendrai pripažystamos anaforų taksonomijos lietuvių kalbai, svarbu įrankyje turėti galimybę laisvai sukurti naujus tipus ar redaguoti esamus.

C. Sprendimo rezultatų analizė – grafinis palyginimas

Kūrėjui, kuriančiam anaforų sprendimo sistemą, svarbu pamatyti konkretių sistemų įverčius. Tokiu būdu kūrėjas gali žinoti savo progresa. Bet tam, kad kūrėjas galėtų tobulinti sistemą, reikia žinoti, kurias anaforas sistema atpažino gerai, o kurių neatpažino ar atpažino klaudingai. Be grafinio rezultatų atvaizdavimo detalių, rankiniu būdu atliekama rezultatų analizė būtų ypač imlius darbui. Rezultatų analizės efektyvumui pagerinti tikslinę naudoti grafinį sprendimo rezultato su etalonu palyginimą, kurį galėtų pateikti kompiuterizuotas įrankis.

D. Duomenų mainai

Pagrindiniai formatai, kuriuos naudoja įvairūs sprendimai [8], [9], [10], veikiantys natūralios kalbos apdorojimo srityje, yra *XML* ir *JSON*. Naudojant *XML* formata, lyginant su *JSON*, labiau yra apkraunama kompiuterio laikinoji atmintis, bet lengviau skaitomas žmogui ir galimas didesnis apdorojimo greitis. Ir atvirkščiai: naudojant *JSON*, reikia mažiau atminties, bet sunkiai žmogaus skaitomas, o programinis apdorojimas lėtesnis nei *XML*. Tieki pirmieji kuriami automatiniai anaforų sprendimai, tiek kiti natūralios kalbos apdorojimo sprendimai lietuvių kalboje (projektas Semantika-LT) duomenims pateikti naudoja *JSON* formatą. Siekiant išlaikyti kitose sistemose priimtų sprendimų tēstinumą ir taip užtikrinti naujų sprendimų integravimą, būtų tikslinę pasirinkti *JSON* formatą. Be to, esant reikalui, yra įmanoma sukurti programas, kurios įgalintų duomenų konversiją tarp formatų.

E. Kiti svarbus aspektai

Lietuvių kalba NKA srityje laikytina neturinti pakankamai išteklių (angl. *under-resourced*), todėl nėra nusistovėjusių poreikių įvertinant lietuvių kalbos anaforų sprendimus. Be to, daugelis įrankių po sukūrimo dažnai yra netobulinami ir nepalaikomi. Todėl atvirojo kodo formatas leistų adaptuoti

įrankių pagal besikeičiančius vartotojų poreikius ir pratęsti įrankio palaikymą ir tobulinimą susiklosčius situacijai, kai autorius nusprendžia nepratęsti įrankio palaikymo ir tobulinimo.

Svarbi įrankio savybe turėtų būti greitas išmokstamumas juo naudotis. Greit išmokus, pavyzdžiu, per dieną, naudotojas galėtų įrankiu dirbtį veiksmingai ir našiai.

F. Esami anaforų anotavimo / įvertinimo sprendimai

Esamiems įrankiams palyginti pasirinkti 8 kriterijai: koreferencijų anotavimo galimybė, anaforų anotavimo galimybė, įvertinimo rezultatų pateikimas matais R, P, F, RR, bendros prieigos galimybė, duomenų mainų palaikymas *JSON* formatu, skirtingu anotacijų grafinio palyginimo galimybė, įrankio tobulinimo galimybė ir ar įmanoma išmokti naudotis įrankiu per 1 dieną. Palyginimui „*Lent. 1*“ buvo pasirinkti 3 įrankiai, kurie tenkina daugiausiai pasirinktų kriterijų:

1) GATE įrankių grupė

GATE [9], [11] įrankiai yra kuriami ir palaikomi ne pelno siekišančios organizacijos, todėl visi grupės komponentai yra atvirojo kodo (sukurti *Java* programavimo kalba). *GATE* yra apdorojimo komponentų integruota kūrimo aplinka (angl. *integrated development environment*) bei informacijos išgavimo (angl. *Information Extraction*) sistema, pateikiama kartu su daug plėtinėlių, leidžiančiu sistemą adaptuoti įvairiems specifiškams poreikiams. Duomenų mainams su išoriniaisiais sprendimais *GATE* naudojamas *XML* formatas. Bet yra sukurtas specifinių *Twitter JSON* palaikantis iškiepis. Standartiniam *JSON* formatui turėtų būti kuriamas papildomas iškiepis. Kalbos specifinės raidės palaikomos naudojant unikodą.

Pagrindinis *GATE* komponentas yra vartotojo kompiuteryje paleidžiama taikomoji programa *GATE Developer*, kurią naudojant galima anotuoti ir grafiškai analizuoti koreferencijų grandinėlę. Taip pat įrankis turi integruotą skirtingu anotacijų įvertinimo funkciją (*GATE AnnotationDiff*), kuri automatiškai apskaičiuoja tikslumą, išsamumą, F-vertę. Bet be papildomų priedų *GATE Developer* negalima žymeti anaforų pirmtakų ir objektų. Jau sukurtų tokų priedų rasti nepavyko. Todėl jie turėtų būti sukurti papildomai. *GATE Teamware* naršykleje veikianti (internetinė) programa ir *GATE Cloud* sudaro bendradarbiavimo galimybes kuriant tekstybus ir jų etalonus.

GATE yra plačiai taikoma sistema, skirta įvairiems NKA naudotojams:

- programuotojams, kuriantiems programinę natūralios kalbos apdorojimo įrangą;
- kalbos tyréjams;
- kalbos apdorojimo mokytojams ir dėstytojams.

Dėl didelio programos pritaikymo galimybių *GATE* išmokti naudotis yra sunku.

2) eHost

eHost [12] įrankis sukurtas Sveikatos apsaugos informaciinių tyrimo konsorciumo (angl. *A Consortium for Healthcare Informatics Research*), i kurį įėjina Jutos universitetas ir Solt Leik Sičio sveikatos apsaugos sistema.

Kūrėjų įrankis, kurio pirminė paskirtis yra medicinos tekstu anotavimas, naudojamas įvairiuose projektuose nuo 2010 m.

Pavadinimas *eHost* yra termino „išplėstinis žmogaus įrankių rinkinys“ (angl. *extensible Human Oracle Suite of Tools*) santrumpa. Tai yra atvirojo kodo *Java* programavimo kalba sukurtas sistemos prototipas, veikiantis per naršykę vietiniame kompiuteryje. *eHost* igalina konceptą, jų savybių ir ryšių tarp jų anotavimą.

Pagrindinės įrankio savybės: tekstu generavimas ir pirminis anotavimas iš žodynų; rankinis ryšių anotavimas; kreipinai į žodynus naudojant adaptuojamą programavimo sąsają (angl. *API*); duomenų mainai *XML* formatu.

3) Anafora

Anafora [10] yra anotavimo įrankis, sukurtas Kolorado universitete. Tai yra atvirojo kodo įrankis, sukurtas *Python* programavimo kalba. Kuriant *Anaforą*, pagrindiniai kriterijai buvo lengvas išmokstamumas ir pritaikumas tiek mažiem, tiek dideliems projektams ir įvairių galimų teksto ryšių anotacijų palaikumas.

Kūrėjų įvardijamas įrankio privalumas ir išskirtinumas – *Anafora* veikia nutolusiai serveryje (debesyje) ir paleidžiama per interneto naršykę. Todėl *Anafora* gali veikti bet kurioje operacinėje sistemoje be papildomo diegimo ar vietinio duomenų saugojimo. Dirbant su programa darbiniai duomenys yra nuolat automatiškai išsaugomi debesyje. *Anafora* duomenų mainams naudoja *XML* formatą.

1 LENTELĖ. ESAMU ĮRANKIŲ PALYGINIMAS

Kriterijus	GATE	eHost	Anafora
Koreferencijų anotavimo galimybė	+	+	+
Anaforų anotavimo galimybė	-	-	-
Įvertinimo rezultatų pateikimas matais R, P, F, RR	+	-	-
Bendros prieigos galimybė	+	-	+
Duomenų mainai <i>JSON</i> formatu	- /+	(<i>Twitter JSON</i>)	-
Skirtingų anotacijų grafinis palyginimas	+	-	-
Įrankio tobulinimo galimybė	+	+	+

Įmanoma išmokti naudotis įrankiu per 1 dieną	-	+	+
--	---	---	---

Iš nagrinėtų įrankių artimiausias pasirinktiems kriterijams yra *GATE*, bet visų keliamų reikalavimų neatitinka. *GATE* yra sudėtinga išmokti naudotis (to padaryti per 1 dieną nepavyko), ir jis yra daugiau skirtas sudėtingiemis kompleksiniams sprendimams kurti, o ne sprendimams analizuoti ir ivertinti. Pagrindinė problema yra ta, kad *GATE* nėra pritaikytas detaliam anaforų analizavimui t. y. nesukūrus papildomų įskiepių neįmanoma atskirai analizuoti anaforos objektą ar pirmaką – galima tik žymėti koreferencijų grandinėles neišskiriant



1 pav. ASAS pagrindinis langas

A. ASAS funkcijos

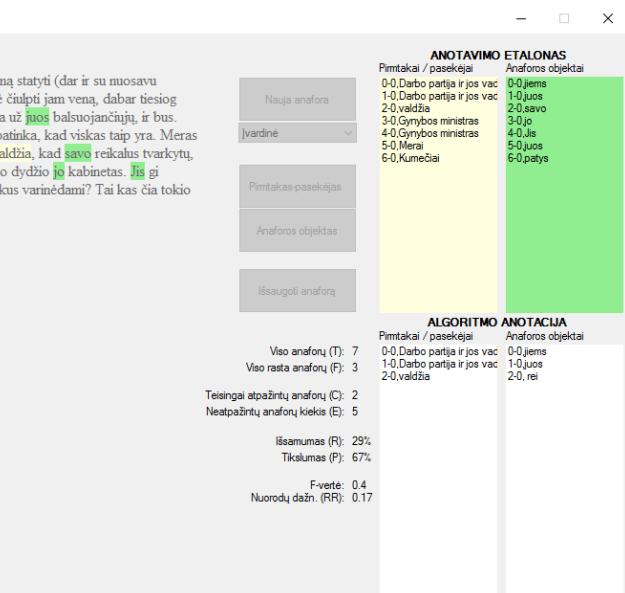
- 1) Prisijungimas – prisijungimas įvedant paskyros duomenis: vartotojo vardą ir slaptažodį;
- 2) Tekstyno sudarymas su galimybėmis:
 - a) nustatyti užbaigtumo žymą tekstyno elementui (tekstui),
 - b) klasifikuoti atskirus tekstyno tekstu pagal žanrą priskiriant jam tipą;
- 3) Etaloninio tekstyno sudarymas, apimantis anaforos objektų ir tipų žymėjimą tekste, parenkant anaforos tipą, su galimybėmis:
 - a) nustatyti užbaigtumo žymą,
 - b) atšaukti užbaigtumo žymą;
- 4) Anaforų sprendimo ivertinimas, apimantis:
 - a) teksto, kuriam sudarytas anaforų etalonas, eksportavimą tekstinio failo pavidalu,

anaforos pirmaką ir objektų. Todėl yra tikslingo sukurti nepriklausomą įrankį, atitinkantį visus reikiamus kriterijus.

III. SPRENDIMAS

Sukurtasis įrankio prototipas ASAS 1 pav. atitinka antrajame skyriuje aprašytus kriterijus, išskyrus skirtingų anotacijų grafinių palyginimą. Įrankis šiuo metu yra tobulinamas ir dar šiais metais (2016 m.) ši funkcija atsiras. Šiame skyriuje pateiktos ASAS realizacijos ir veikimo detalės.

ASAS veikia vartotojo kompiuteryje su Windows operacine sistema ir suteikia jos naudotojui galimybę sudaryti anaforų anotavimo etalonus ir juos palyginti su anaforų sprendimo sistemos rezultatais, išvesdamas ivertinimo matus.



b) teksto etaloninės anaforų anotacijos eksportavimą JSON formatu tekstinio failo pavidalu,

c) anaforų sprendimo sistemos rezultatų, pateiktų tekstiniaiame faile JSON formatu, importavimą,

d) ivertinimo rezultatų gavimą – grafinį atvaizdavimą ir įverčių R, P, F, RR pateikimą;

5) ASAS parametru konfigūravimas, apimantis:

a) naujų anaforų tipų ir jų kodų sukūrimą ir senų tipų (kodų) redagavimą,

b) naujų teksto tipų sukūrimą ir senų tipų redagavimą, c) užbaigtumo žymos atšaukimą.

B. Sprendimo igyvendinimas

Prototipas ASAS suprogramuotas programavimo kalba C#. Įrankis interne prieinamas adresu <http://asas.netseptyni.lt>

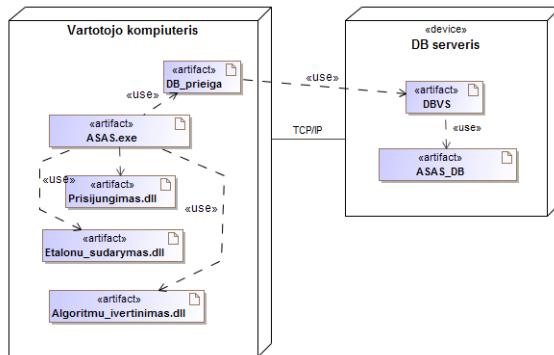
ASAS yra išleistas Microsoft OneClick Deployment technologija. Naudojama Microsoft SQL Server duomenų bazė. Įrankio duomenų bazė ir diegimo prieiga veikia tarnybinėje stotyje, prie kurios jungiamasi su kliento dalimi, kuri yra

įdiegama vartotojo kompiuteryje iš diegimo prieigos 2 pav. Todėl ASAS kliento dalis gali veikti tik tuomet, kai yra ryšys su tarnybinės stoties ASAS dalimi. ASAS, kiekvieną kartą prieš pradėdamas darbą (t. y. kai paleidžiama kliento dalis vartotojo kompiuteryje), tikrina ar nėra išleistas atnaujinimas. Jei yra – automatiškai atsinaujina.

Ši diegimo schema suteikia daug privalumų:

- Serverio dalies apkrova: apkraunamas tik *SQL* serveris, nes visa programos logika vykdoma vartotoju kompiuteriuose. Tai leidžia surūpinti resursų techninei įrangai;
- Jokie duomenys nėra saugomi vartotojų kompiuteriuose (debesų technologija), todėl jie yra vientisi ir visiems prieinami vienodai. Vartotojams nereikia rūpintis dėl atsarginių duomenų kopijų. Naujiems vartotojams prieiga prie duomenų gali būti suteikiama paprasčiausiai sukuriant jiems paskyrą, įgalinančią naudotis įrankiu;
- Ypač paprastas įrankio diegimas, atliekamas vienu diegimo failo paspaudimu jį paleidžiant;

Šios technologijos pagrindinis trūkumas yra tai, kad įrankis gali būti įdiegtas tik *Windows* operacinėje sistemoje su įdiegtu ne žemesniu nei 4.5 versijos *.Net* karkasu.



2 pav. ASAS diegimo diagrama

C. Duomenų modelis

Duomenų mainams ir saugojimui naudojama biblioteka *JSON.Net 8.0.2*. Anotacijos yra išrašomos *JSON* formatu duomenų bazėje *string* tipo išraše. Pats tekstas saugomas atskiru *string* išrašu. Sukūrus kiekvieną tekštą, jam yra priskiriamas unikalus identifikacinis numeris, kuris *JSON* turi būti pateikiamas priekyje. ASAS ši numerį naudoja teksto identifikavimui importuojant anotaciją. Tiek anaforų pirmtakai, tiek objektai yra saugomi dviečiems sveikais skaičiais, kurių pirmasis nusako pradžios simbolį tekste, o antrasis ilgi. Anaforos tipo kodo išsaugojimas šiuo metu neigyvendintas, nes nėra anaforų kodų klasifikacijos. Kadangi kodui vieta numatyta iš anksto, jo vietoje yra saugoma *null* reikšmė. 3 pav. pateiktas teksto ir jo anotacijos grafinės vizualizacijos ASAS programeje ir anotacijos *JSON* formatu tekstiniame faile, pavyzdys.

Tomas nejo į mokyklą. Jis sirgo. Mokytoja apie tai informuota nebuvo, todėl ji sunerimo. Vaikai visada privalo informuoti el. paštu.

Tomas nejo į mokyklą. Jis sirgo. Mokytoja apie tai informuota nebuvo, todėl ji sunerimo. Vaikai visada privalo informuoti el. paštu.

{"teksto_id": "1014", "anaforos": [{"pirmtakai": ["0,5"], "objektai": ["22,3"], "tipo_kodas": null}, {"pirmtakai": ["33,8"], "objektai": ["76,2"], "tipo_kodas": null}]}]

3 pav. Pavyzdinis tekstas ir jo grafinė bei JSON anotacijos.

D. ASAS veikimo aprašymas ir pagrindinių funkcijų atlikimo scenarijus

ASAS prieš pateikdamas grafinę sąsają vartotojui, pirma automatiškai jungiasi prie diegimo tarnybinės stoties ir patikrina, ar nėra išleistų atnaujinimų.

Jei ASAS negali užmegzti ryšio su diegimo ir duomenų serveriu, ASAS paleidimas yra nutraukiamas ir parodos klaidos pranešimas. Užmezgus ryšį ir radus atnaujinimų – ASAS automatiškai atsinaujina naujausių leidimų. Atsinaujinus ar neradus atnaujinimų, atidaromas autorizacijos langas, kuriame reikia įvesti vartotojo vardą ir slaptažodį. Vartotojas autorizuojasi savo paskyra ir pradedą darbą. Paskyros tipas gali būti dviejų tipų: tyrėjas arba administratorius. Nuo to priklauso prieinamų funkcijų sąrašas. Tyrėjo role dirbantis vartotojas gali atlkti 1, 2, 3, 4 funkcijas. Administratorius – visas 5 funkcijas, t. y. gali konfigūruoti įrankio ASAS parametrus.

I. Funkcijos „2. Tekstyno sudarymas“ scenarijus

- Vartotojas turi būti prisijungęs tyrojo arba administratoriaus rolo;
- Pagrindiniam lange pasirenkama „Sukurti naują tekštą“ arba „Įkelti esamą tekštą iš tekstyno“;
- Pasirinkus naujo teksto kūrimą, rodomas tuščias teksto įvedimo langas ir kategorijos priskyrimo meniu su neparinkta kategorija. Įkėlus jau esamą tekštą, lange rodomas įkeltas teksts, o menu – esama priskirta kategorija. Tol, kol tekstu nėra uždėta žyma „baigtas“, laikoma, kad teksts yra nesukurtas ir dar gali būti redaguojamas: papildomas, taisomas, keičiama kategorija, bet negali būti pradedamas anaforų žymėjimas (anotavimo etalonu sudarymas);
- Tekstas papildytas gali būti dviečių būdu: įvedant tekštą rankiniu būdu arba įklijuojant klavišų kombinaciją *Ctrl + V*;
- Sudarant naują tekštą arba, toliau tēsiant jau sukurto teksto redagavimą, tekstas nėra išsaugomas tol, kol nėra pasirenkama išsaugoti. Tekštą išsaugant pirmą kartą (kuriant naują tekštą), privalo būti nustatyta teksto kategorija. Pirmo išsaugojimo metu tekstas yra pridedamas į tekstyntą. Testinio redagavimo metu, teksto kategorija gali būti pakeista.

II. Funkcijos „3. Etaloninio tekstyno sudarymas“ scenarijus

- Vartotojas turi būti prisijungęs tyrojo arba administratoriaus rolo;
- Pagrindiniam lange pasirenkama „Įkelti esamą tekštą iš tekstyno“;

- Tuomet pagrindiniame lange atsiranda galimybė įkelti etaloną, jei tekstas turi žymą „baigtas“.
- Užkrovus etaloną, sarašuose rodomas tekste sužymėtos anaforos. Anaforos taip pat yra pateikiamos ir grafiškai skirtingomis spalvomis atvaizduojant pirmakus ir objektus;
- Tol, kol etalonui nėra priskirta žyma „baigtas“, etalone sudarymas gali būti pratęstas: trinamos jau pažymėtos anaforos ar pažymimos naujos;
- Sudarant etaloną, anaforų žymėjimas ar korekcijos nėra išsaugomos tol, kol neparenkama išsaugoti.

III. Funkcijos „4. Anaforų sprendimo įvertinimas“ scenarijus

- Vartotojas turi būti prisijungęs tyrejo arba administratoriaus role;
- Pagrindiniame lange pasirenkama užkrauti esamą tekstą iš tekstyno;
- Tuomet pagrindiniame lange atsiranda galimybė užkrauti etaloną, jei tekstas turi žymą „baigtas“;
- Jei etalonas turi žymą „baigtas“, galima eksportuoti neanotuotą tekstą tekstinio failo pavidalu. Eksportuotas tekstas skirtas anaforų sprendimo sistemai anotuoti arba eksportuoti jį kaip etaloninę anotaciją;
- Anaforų sprendimo įvertinimas atliekamas importuojant anaforų sprendimo anotaciją. Importavus automatiškai yra išvedami įvertinimo kriterijai. Laikoma, kad anafora sprendimo buvo atpažinta, jei buvo atpažinti visi pirmtakai ir visi objektais nesuklystant nei vienu simboliu.

IV. Funkcijos „5. ASAS parametrų konfigūravimas“ scenarijus

- Vartotojas turi būti prisijungęs tyrejo arba administratoriaus role;
- Pasirenkamas nustatymų meniu;
- Nustatymų meniu galima pridėti, ištinti ar redaguoti esamus anaforų tipus, anaforų kodus ir tekstu kategorijas 4 pav.

Nustatymai

Anaforų tipai:		Tekstų kategorijos:	
Tipų kodas	Tipo pavadinimas	Kategorija	
A001	Ivardinė	Medicina	
		Naujienos	
		Politika	
		Teisė	

4 pav. Nustatymų langas

IŠVADOS IR ATEITIES DARBAI

Šiame straipsnyje buvo pristatyta lietuvių kalbos anaforų sprendimo analizės ir įvertinimo prototipas ASAS. ASAS šiuo metu, kiek žinoma, yra vienintelis tam skirtas lietuviškas įrankis.

Atlikta esamu įrankių analizė parodė, kad nei vienas iš jų neatitiko visų iškeltų reikalavimų, todėl buvo nuspresta sukurti

naują nepriklausomą prototipą. Sukurtas prototipas atitinka 7 iš 8 jam iškeltų pagrindinių reikalavimų, bet šiuo metu yra tobulinamas ir šiais metais (2016 m.) bus užbaigtas.

Kadangi šiuo metu (2016 m. vasario mėn.) lietuvių kalbos anaforų anotavimo tekstynei ir jų etalonai, naudojanties ASAS, yra tik sudarinėjami, nebuvu atliktas prototipo eksperimentas su didesniais duomenų kiekiuais. Dėl to nėra aiškus ASAS veikimo stabilumas su didelėmis duomenų apimtimis. Su mažais duomenų kiekiuais ASAS veikė be klaidų ir gavo teisingus imitacinio anaforų sprendimo įvertinimo matus.

Šiuo metu KTU Informacijos sistemų katedroje yra kuriami automatiniai anaforų sprendimai (projekto Semantika-LT tasa), kurių analizė galės būti atliekama naudojant ASĄ.

Tolimesniais darbais bus siekiama sudaryti automatinio anaforų sprendimo įvertinimo metodiką ir tobulinti jau sukurtą įvertinimo ir analizės prototipą.

LITERATŪRA

- [1] R. Mitkov, „Anaphora Resolution“ London, UK, 2002.
- [2] A. Čiukšys ir L. Nemuraitė, „Ko-referencijų sprendimo įrankio sukūrimo lietuvių kalbai galimybių analizė,“ įtraukta 20-oji tarpuniversitetinė magistrantų ir doktorantų konferencija: „Informacinių technologijos 2015“, Kaunas, 2015.
- [3] N. Chincor, D. Lewis ir L. Hirschman, „Evaluating Message Understanding Systems: An Analysis Of The Third Message Understanding Conference (MUC-3)“ *Computational Linguistics Journal*, 1993.
- [4] R. Grishman ir B. Sundheim, „Message Understanding Conference-6: a brief history,“ *Computational linguistics*, t. 1, 1996.
- [5] H. Lee, A. Chang, Y. Peirsman, N. Chambers, M. Surdeanu ir D. Jurafsky, „Deterministic Coreference Resolution Based on Entity-Centric, Precision-Ranked Rules,“ *Computational Linguistics*, 1 t. iš 239-4, pp. 886-916, 2013.
- [6] B. Xian, F. Zahari ir D. Lukose, „Benchmarking ARS: anaphora resolution system,“ įtraukta 11th International Conference on Knowledge Management and Knowledge Technologies, 2011.
- [7] V. Žitkus ir L. Nemuraitė, „Taxonomy of anaphoric expressions as a starting point for anaphora resolution in Lithuanian corpus,“ įtraukta 19-oji tarpuniversitetinė magistrantų ir doktorantų konferencija: IVUS, Kaunas, 2014.
- [8] C. Muller ir M. Strube, „MMAX: A Tool for the Annotation of Multi-modal Corpora,“ įtraukta Proceedings of the 2nd IJCAI Workshop on Knowledge and Reasoning in Practical Dialogue Systems, 2001.
- [9] The University of Sheffield, Department of Computer Science, „GATE,“ [Tinkle]. Available: <https://gate.ac.uk/>. [Žiūrėta 2016 02 18].
- [10] W. Chen ir W. Styler, „Anafora: A Web-based General Purpose Annotation Tool,“ įtraukta Proceedings of the NAACL HLT 2013 Demonstration Session, Atlanta, Georgia, USA, 2013.
- [11] H. Cunningham, D. Maynard, K. Bontcheva ir V. Tablan, „GATE: an Architecture for Development of Robust HLT Applications,“ įtraukta Proceedings of 40th Annual Meeting of the Association for Computational Linguistics, 2002.
- [12] C. Leng, B. South ir S. Shuying, „eHOST: The Extensible Human Oracle Suite of Tools | BLUlab,“ [Tinkle]. Prieinama: <http://blulab.chpc.utah.edu/content/ehost-extensible-human-oracle-suite-tools/>. [Žiūrėta 2016 02 18].

New Radial Basis Function Method Modelling Borehole Heat Transfer: the theoretical survey

Audrius Indriulionis, Juozas Bielskus and Pranas Vaitkus

Abstract—The borehole heat exchangers (BHE) are used to transfer heat from/to the ground or heat could be stored in the ground for the specific needs. During the last two decades the various analytical and numerical methods were developed to predict and analyse in order to design and optimize the BHE installation and performance parameters. In general, the heat transfer in BHE is divided into three parts: inside BHE, around the BHE, and the interactions between BHE's including the long-time and short-time heat transfer analysis. In this work, the method of radial basis function (RBF) is used for the borehole heat transfers problems. Because a much wider range of physical phenomena are modelled by the RBF method is an efficient mesh free technique for the numerical solution of partial differential equations. The main advantage of numerical methods which use radial basis functions over traditional techniques is the meshless property of these methods. In a meshless method, a set of scattered nodes are used instead of meshing the domain of the problem. Solving the system of linear equations usually does not lead to accurate results by most numerical methods, because condition number of matrix A is large. It means that the ill-conditioning of matrix A makes the numerical solution unstable. Tichonov regularization (TR) method are presented to solve such ill-conditioned systems. In this work, generalized cross-validation (GCV) method is carried out to determine the regularization parameter ξ for the TR method that minimizes the GCV function. The results of numerical experiments are presented and some comparisons are made with well-known finite elements schemes.

Index Terms—borehole heat exchangers (BHE), radial basis function (RBF), Tichonov regularization (TR), generalized cross-validation (GCV).

I. INTRODUCTION

THE one of the first mathematical books for heat conduction in solids was first time introduced by Carslaw and Jager [1], [2]. The PDE (partial differential equation) of heat conduction was solved under the general Dirichlet and Neuman initial initial and boundary conditions. Since 1990's the numerous applications for borehole heat exchangers have been developed. The BHE models are divided into analytical and numerical models. Analytical solutions such as

A. Indriulionis PhD Student, Department of Quaternary Research, Institute of Geology and Geography, Nature Research Center, Akademijos str. 2, LT-08412 Vilnius, Lithuania, e-mail:audrius.indriulionis@gmail.com.

J. Bielskus PhD Student, Department of Building Energetics, Faculty of Environmental Engineering, Vilnius Gediminas Technical University, Saulėtekio av. 12, LT-10223 Vilnius, Lithuania, e-mail:juozas.bielskus@vgtu.lt.

Associate professor dr. Pranas Vaitkus, Department of Mathematical Statistics, Faculty of Mathematics and Informatics, Vilnius University, Naugarduko str. 24, LT-03225 Vilnius, Lithuania, e-mail:pranas.vaitkus@mif.vu.lt.

line-source [3], [4], [5], cylinder-source [1], [6], [7] , [8], [9], [10], [11] have been used for analysis and dimensioning of VBHE. The numerical models are more accurate than analytical models especially on short-time scales, but may be computationally inefficient incorporating directly into design and energy analysis program. The simulated data of large number of complex grids are pre-computed and stored in huge DB [12]. The traditional numerical techniques have been used by various numerical methods such as finite differences, finite volumes, finite elements etc. by [13], [14], [15], [16], [17], [18] and [19] to model VBHE. The numerical schema were based on discretization using polar or cylindrical grids or meshes. Especially for multidimensional problems formulated on complexly shaped domains, mesh generation can be very time consuming. Recently, meshless methods for the numerical solution of PDEs became popular between scientists and engineers. In many geological situations when dealing with a heat conduction in the ground it is not always possible to specify the boundary conditions or the initial ground temperature. However in many heat transfer situations, the surface heat flux and temperature histories must be determined from transient temperature measurements at one or more interior locations. This is an inverse heat conduction problem. Briefly speaking one might say the inverse problems are concerned with determining causes for a desired or an observed effect. The main difficulty in the treatment of inverse problems is the instability of their solution in the presence of noise in the measured data. According to the fact that unknown solutions of inverse problems are determined through indirect observable data which contain measurement errors, such problems are naturally unstable. From another point of view, since the existence, uniqueness, and stability of the solutions of these problems are not usually confirmed, they are generally identified as ill-posed, [20] [21], [22]. Hence, several numerical methods have been proposed for solving the various kinds of inverse problems. In addition to, ill-posedness of these kinds of problems, ill-conditioning of the resulting discretized matrix is the main problem making all numerical algorithms for determining the solution of these kinds of problems. Accordingly, within recent years, meshless methods, as the method of fundamental solution (MFS), radial basis functions (RBFs) method, and some other methods, have been applied by many scientists in the field of applied sciences and engineering. We will present a meshless numerical scheme, based on the radial basis function of the heat equation, in order to approximate the solution of a backward inverse

heat conduction problem (BIHCP), the problem in which an unknown initial condition or/and temperature distribution in previous time will be determined. This kind of problem may emerge in many practical application areas such as geology. On the other hand, since the system of the linear equations obtained from discretizing the problem in the presented method is ill-conditioned, Tikhonov regularization (TR) method is applied in order to solve it. The generalized cross-validation (GCV) criterion has been assigned to adopt an optimum amount of the regularization parameter. The structure of the rest of this work is organized as follows: In Section II, we represent the mathematical formulation of the problem. The Numerical study using Radial Basis Function method (RBF) are described in Section III. Section IV embraces Tikhonov regularization method with a rule for choosing an appropriate regularization parameter. In Section V, we present the obtained numerical results of solving a problem. Section VI ends in a brief conclusion and some suggestions.

II. MATHEMATICAL FORMULATION OF THE PROBLEM

Let's denote $\mathbf{x} = (x, y)$ two-dimensional case where r^* could be the distance values from the interval $[r_{p-}^*, R]$. We define for different material layers and denote it as $k = f$ (for fluid), p (for pipe), g (for grout), s (for soil), b (for borehole), special notation $-$, $+$ would be used for the left and right side of the boundary. The radial heat transfer model is considered approximating the U-tube on an equivalent-diameter pipe. The general input parameters of the model are: q_{inj} - a constant heat flux, a fluid thermal capacity C_f of the circulating fluid, the S_f intersection area of the circulating and in the equivalent-diameter pipe is kept equal to the U-tube, a R_p is accounted for both fluid and pipe resistances. The circulating fluid at temperature $u_{fluid}(\tau)$ starting from time $\tau = 0$ changing the temperatures $u_g(r_{p+}^*, \tau)$ and $u(r_{p+}^*, \tau)$ on outer boundary of the equivalent-diameter pipe in the grout region and at the borehole radius r_{b+}^* on the surrounding soil, accordingly. Thermal conductivity and thermal diffusivity parameters could be described on II. A grout region of thermal conductivity λ_g and thermal diffusivity α_g surrounds the equivalent-diameter pipe. The borehole is surrounded by infinite homogeneous soil of thermal conductivity $\lambda_s(r^* > r_{b+}^*)$ and thermal diffusivity $\alpha_s(r^* > r_{b+}^*)$. The q_{inj} heat flux from the fluid to the grout region through the pipe wall is $q(r_{p+}^*, \tau)$. Similarly, the heat flux from the grout region to the surrounding soil $q(r_{b+}^*, \tau)$, where the borehole radius is r_{b+}^* . The resulting radial heat transfer problem is shown in Figure 1 carefully changing the general schema from [11]. For the heat transfer problem shown in figure 1, the temperature distribution $u(r^*, \tau)$ must satisfy the following radial heat conduction equation in both the grout and the ground (soil) regions.

$$\frac{1}{\alpha_k(r^*)} \cdot \frac{\partial u_k}{\partial \tau} = \frac{\partial^2 u_k}{\partial^2 r^*} + \frac{1}{r^*} \cdot \frac{\partial u_k}{\partial r^*}, \quad (1)$$

where $k = g, s$. The radial heat flux in the grout and the soil regions is

$$q(r^*, \tau) = 2\pi r^* (-\lambda(r^*)) \frac{\partial u_k}{\partial r^*}, \quad (2)$$

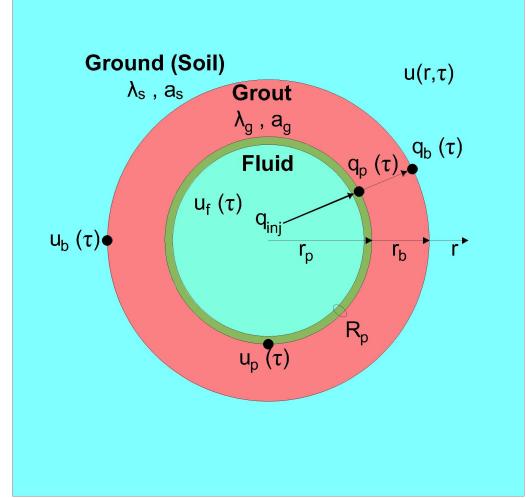


Fig. 1. Borehole heat transfer process changed from [11]

TABLE I
SOME RADIAL BASIS FUNCTIONS

RBF	Definition
Multiquadratic(MQ)	$\phi(\mathbf{x}, \mathbf{x}_j) = \phi(r_j) = \sqrt{r_j^2 + \epsilon_j^2}$
Inverse Multiquadratic(IMQ)	$\phi(\mathbf{x}, \mathbf{x}_j) = \phi(r_j) = \sqrt{r_j^2 + \epsilon_j^2}^{-1}$
Inverse Quadratic(IQ)	$\phi(\mathbf{x}, \mathbf{x}_j) = \phi(r_j) = (r_j^2 + \epsilon_j^2)^{-1}$
Gaussian (GA)	$\phi(\mathbf{x}, \mathbf{x}_j) = \phi(r_j) = \exp^{-(r_j^2 + \epsilon_j^2)}$
Bassell RBF	$\phi(\mathbf{x}, \mathbf{x}_j) = \phi(r_j) = J_0(k(\sqrt{r_j^2 + \epsilon_j^2}))$

where $k = g, s$. The heat flux at the grout-soil interface is continuous and the boundary condition $r^* = r_{b+}^*$ is

$$\lambda_{grout} \cdot \frac{\partial u}{\partial r^*} \Big|_{r^* = r_{b+}^*} = \lambda_{soil} \cdot \frac{\partial u}{\partial r^*} \Big|_{r^* = r_{b+}^*}, \quad u(R, \tau) = const. \quad (3)$$

The boundary condition at the pipe-grout interface then is

$$u(r_{p-}^*, \tau) - u(r_{p+}^*, \tau) = R_p \cdot q(r_{p+}^*, \tau). \quad (4)$$

Here the thermal resistance R_p is for both fluid and pipe is defined as

$$R_p = \frac{1}{2\pi r_{p+}^* h_p} + \frac{1}{2\pi \lambda_{pipe}} \ln \left(\frac{r_{p+}^*}{r_{p-}^*} \right). \quad (5)$$

The heat balance of the fluid in the pipe with the injected heat q_{inj} is

$$q_{inj} = C_f S_f \frac{\partial u(r_{p-}^*, \tau)}{\partial \tau} + q(r_{p+}^*, \tau). \quad (6)$$

The initial temperatures of the fluid in the pipe, the grout and the soil regions are taken as constants.

$$u(r_{p-}^*, 0) = const., \quad u(r^*, 0) = const. \quad r^* \geq r_{p+}^*. \quad (7)$$

III. NUMERICAL STUDY

A. Radial Basis Function

The $\phi_j(\mathbf{x}) = \phi(r)$ and $\mathbf{x} = (x, y)$ where $r = \|\mathbf{x} - \mathbf{x}_j\|_2$ be the Euclidean distance between \mathbf{x}_j and \mathbf{x} are fixed point and arbitrary point in \mathbb{R}^d , respectively. RBF's ϕ are radially symmetric about \mathbf{x}_j e.g. depends on distance between \mathbf{x}_j and \mathbf{x} . Some well known infinitely smooth RBF's are listen in Table III, where ϵ the empirically chosen shape parameter who plays an important role using RBF. The main idea of using the RBF is the approximation of solution $u(\mathbf{x})$ which could be expressed with linear combination of RBF Table III.

B. Approximate Solution using RBF

$$u^n(\mathbf{x}) = u(\mathbf{x}, n \Delta \tau) = \sum_{j=1}^N c_j \phi(\mathbf{x}, \mathbf{x}_j), \quad (8)$$

where c_j a unknown coefficients for $j = 1, \dots, N$, to be determined by collocation technique. In the case of RBF method, the spatial domain and its boundary are represented by a set of scattered nodes (they coincide with the centers of RBF's). $\bar{\Omega} = \Omega \cap \Gamma$, where $\Omega = \Omega_{r_p^*} \cap \Omega_R$ and $\Gamma = \Gamma_{r_p^*} \cap \Gamma_{r_b} \cap \Gamma_R$. Let denote the set of collocation points(nodes) $\Xi = \{(x_j, y_j)\}_{j=1}^N \subset [r_p^*, R] \times [0, \tau_{max}]$, where the total number of nodes in each domain is equal to

$$N = N_{\Gamma_{r_p^*}} + N_{\Gamma_{r_b^*}} + N_{\Gamma_R} + N_{\Omega_{r_b}} + N_{\Omega_R}.$$

We assume the $\Xi = \Xi_{\Gamma_{r_p^*}} \cap \Xi_{\Gamma_{r_b^*}} \cap \Xi_{\Gamma_R} \cap \Xi_{\Omega_{r_b^*}} \cap \Xi_{\Omega_R}$, and r_p^*, r_b^*, R are known constants where

$$\Xi_{\Gamma_\tau} = \{(x_j, y_j) \in \Xi : r_{p-}^* \leq \sqrt{(x_j^2 + y_j^2)} \leq R, \quad \forall \tau = 0\}$$

$$\Xi_{\Gamma_{r_p^*}} = \{(x_j, y_j) \in \Xi : \sqrt{x_j^2 + y_j^2} = r_{p-}^*, \quad \forall \tau\}$$

$$\Xi_{\Gamma_{r_b^*}} = \{(x_j, y_j) \in \Xi : \sqrt{x_j^2 + y_j^2} = r_b^*, \quad \forall \tau\}$$

$$\Xi_{\Gamma_R} = \{(x_j, y_j) \in \Xi : \sqrt{x_j^2 + y_j^2} = R, \quad \forall \tau\}$$

$$\Xi_{\Omega_{r_b^*}} = \{(x_j, y_j) \in \Xi : r_{p+}^* \leq \sqrt{x_j^2 + y_j^2} \leq r_{b-}^*, \quad \forall \tau\}$$

$$\Xi_{\Omega_R} = \{(x_j, y_j) \in \Xi : r_{b+}^* \leq \sqrt{x_j^2 + y_j^2} < R, \quad \forall \tau\}$$

and $\Xi_l \neq \emptyset$, $l = \Gamma_{r_p^*}, \Gamma_{r_b^*}, \Gamma_R, \Omega_{r_b^*}, \Omega_R, \Gamma_\tau$.

C. Discretization Schema

The shape parameter ϵ of RBF could be found by numerical experiments and the coefficients c_j^0 are determined the following equation 7

$$\sum_{j=1}^N c_j^0 \phi(\mathbf{x}_i, \mathbf{x}_j) = u_0(\mathbf{x}_i), \quad i = 1, \dots, N, \quad \mathbf{x}_i \in \Xi_{\Gamma_\tau} \quad (9)$$

Initially, it is assumed that $c_j^n = c_j^{n-1}$ and the equations below can be solved for c_j^n . For internal domains the system of

equations should be used the Cranck - Nicolson schema with $\theta = \frac{1}{2}$ and

$$\begin{aligned} \frac{1}{\alpha(r^*) \Delta \tau} u^n(\mathbf{x}_i) - \theta \sum_{j=1}^N c_j^n \frac{\partial^2 \phi(\mathbf{x}_i, \mathbf{x}_j)}{\partial \mathbf{x}_i^2} = \\ \frac{1}{\alpha(r^*) \Delta \tau} u^{n-1}(\mathbf{x}_i) - (1-\theta) \sum_{j=1}^N c_j^{n-1} \frac{\partial^2 \phi(\mathbf{x}_i, \mathbf{x}_j)}{\partial \mathbf{x}_i^2}, \quad (10) \\ i = 1, \dots, N, \quad \mathbf{x}_i \in \Xi_{\Omega_{r_b^*}} \cup \Xi_{\Omega_R} \end{aligned}$$

The system of equations describe the boundary conditions below

$$\sum_{j=1}^N c_j^n \phi(\mathbf{x}_i, \mathbf{x}_j) = u_0(R, n \Delta \tau), \quad i = 1, \dots, N, \quad \mathbf{x}_i \in \Xi_{\Gamma_R} \quad (11)$$

$$\sum_{j=1}^N c_j^n \frac{\partial \phi(\mathbf{x}, \mathbf{x}_j)}{\partial \mathbf{x}} \Big|_{\mathbf{x}=r_{p+}^*} = \frac{q_{r_{p+}^*}}{2\pi r_{p+}^* (-\lambda(r^*))}, \quad (12) \\ r_{p+}^* \in \Xi_{\Gamma_{r_{p+}^*}}$$

$$\lambda_s \sum_{j=1}^N \frac{\partial c_j^n \phi(\mathbf{x}, \mathbf{x}_j)}{\partial \mathbf{x}} \Big|_{\mathbf{x}=r_{b+}^*} = \lambda_g \sum_{j=1}^N \frac{\partial c_j^n \phi(\mathbf{x}, \mathbf{x}_j)}{\partial \mathbf{x}} \Big|_{\mathbf{x}=r_{b-}^*}, \\ r_b^* \in \Xi_{\Omega_{r_b^*}} \cup \Xi_{\Omega_R} \quad (13)$$

$$\left(1 + \frac{C_f S_f R_p}{\Delta \tau}\right) \sum_{j=1}^N c_j^n \phi(r_{p-}^*, \mathbf{x}_j) - \sum_{j=1}^N c_j^n \phi(r_{p+}^*, \mathbf{x}_j) = \\ R_p \cdot q_{inj} + \frac{C_f S_f R_p}{\Delta \tau} \cdot \sum_{j=1}^N c_j^{n-1} \phi(r_{p+}^*, \mathbf{x}_j), \\ r_{p-}^* \in \Xi_{\Gamma_{r_p^*}}, r_{p+}^* \in \Xi_{\Omega_{r_b^*}} \quad (14)$$

This system of equations can be written in the following matrix form as

$$\mathbf{A} \mathbf{c}^{(n)} = \mathbf{b} \quad (15)$$

where $\mathbf{c} = (\mathbf{c}_1^n, \mathbf{c}_2^n, \dots, \mathbf{c}_N^n) \quad \forall n \in (0, \tau_{max})$ and coefficients matrix \mathbf{A} a_{ij} are often ill-conditioned [21], [23] for both direct and inverse [20] heat transfer problems. The prediction of approximate solution $u(\mathbf{x}, n \Delta \tau)$ would be unstable especially than the input data contains noise. The shape parameter ϵ also controls the accuracy of the method. In order to get the accurate results the small value of shape parameter ϵ is required. We assure that the matrix \mathbf{A} doesn't tend to be very ill-conditioned because of $\epsilon \rightarrow 0$. At the same time the method should be very accurate and not well-conditioned the uncertainty principle should be satisfied.

IV. REGULARIZATION METHOD AND PARAMETER

Before presenting the regularization method and regularization parameter, we introduce the singular value decomposition (SVD) of the coefficient matrix \mathbf{A} ,

$$\mathbf{A} = \mathbf{W}\Sigma\mathbf{V}^T, \quad (16)$$

where $\mathbf{W} = [w_1, w_2, \dots, w_N] \in \mathbb{R}^{N \times N}$, $\mathbf{W}^T\mathbf{W} = \mathbf{I}_N$ and $\mathbf{V} = [\nu_1, \nu_2, \dots, \nu_N] \in \mathbb{R}^{N \times N}$, $\mathbf{V}^T\mathbf{V} = \mathbf{I}_N$, where \mathbf{I}_N denotes the N -th order identity matrix. The singular values of matrix \mathbf{A} are the diagonal entries of $\Sigma = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_N)$ which are non-negative and are arranged in increasing order $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_N \geq 0$. Using SVD it's easy to get solution for the

$$\mathbf{c} = \sum_{j=1}^N \frac{w_j^T \mathbf{b} \nu_j}{\sigma_j}. \quad (17)$$

The conventional L_2 condition of \mathbf{A} is defined as $\text{Cond}(\mathbf{A}) = \frac{\sigma_1}{\sigma_N}$, in which σ_1 and σ_N are the largest and smallest singular values of \mathbf{A} .

A. Regularization method

Solving the system of linear equations usually does not lead to accurate results by most numerical methods, because condition number of matrix \mathbf{A} is large. It means that the ill-conditioning of matrix \mathbf{A} makes the numerical solution unstable. Now, in methods as the proposed method in this study, which is based on RBFs, the condition number of \mathbf{A} depends on some factors such as the shape parameter ϵ , as well. On the other hand, for fixed values of the shape parameter ϵ , the condition number increases with the number of scattered nodes N . In practice [24] Matlab toolbox was used, the shape parameter must be adjusted with the number of interpolating points. Also, the accuracy of radial basis functions relies on the shape parameter. So, in case a suitable amount of it is chosen, the accuracy of the approximate solution will be increased. Despite various research works which are done, finding the optimal choice of the shape parameter is still an open problem. Accordingly, some regularization methods are presented to solve such ill-conditioned systems. Tikhonov regularization (TR) method is mostly used by researchers. Since, small perturbation in initial data may produce a large amount of perturbation in the solution, we use the **rand** function in Matlab in the numerical example presented and we produce noisy data as the follows:

$$b^* = b_i(1 + \delta \text{rand}(\mathbf{i})), i = 1, \dots, N$$

where b_i is the exact data, $\text{rand}(\mathbf{i})$ is a random number uniformly distributed in $[1, 1]$, and the magnitude δ displays the noise level of the measurement data. In this method, the Tichonov regularized solution \mathbf{c}^ξ for the system of linear equations is explained as the solution of the following minimization problem:

$$\min(\|\mathbf{Ac} - \mathbf{b}\|_2^2 + \xi^2 \|I_N \mathbf{c}\|_2^2) \quad (18)$$

where $\|\cdot\|_2$ denotes the Euclidean norm and ξ is called the regularization parameter.

B. Regularization parameter

In this work, we apply GCV [25] to obtain regularization parameter ξ . The optimal value of regularization parameter ξ minimizes the following (GCV) function:

$$GCV(\mathbf{c}) = \frac{\|\mathbf{Ac}^\xi - \mathbf{b}\|_2^2}{\text{trace}(\mathbf{I}_N - \mathbf{AA}^T)^2}$$

where $\mathbf{A}^T = (\mathbf{A}^T \mathbf{A} + \xi^2 \mathbf{I}_N)^{-1} \mathbf{A}^T$. The regularization parameter ξ have to use for the indirect problems and GCV procedure used to solve such ill-conditioned or ill-posed problems. The regularized solution is shown by $\mathbf{c}^\xi = [\mathbf{c}_1^\xi, \mathbf{c}_2^\xi, \dots, \mathbf{c}_N^\xi]$, ξ is a minimizer of GCV. The approximate solution for mathematical problem is written as following

$$u^*(\mathbf{x}) = u(\mathbf{x}, n \Delta \tau) = \sum_{j=1}^N c_j^\xi \phi(\mathbf{x}, \mathbf{x}_j) \quad (19)$$

V. NUMERICAL EXPERIMENT

In this section the numerical results were validated and investigated the performance of RBF method. The exact (numerical) solution was implemented using the Comsol Multiphysics software applying the finite elements method. We define the root mean square error (RMSE), relative root mean square error (RRMSE), the maximum absolut error E_∞

$$\begin{aligned} RMSE(u) &= \sqrt{\frac{1}{N_t} \sum_{k=1}^{N_t} (u(\mathbf{x}) - u^*(\mathbf{x}))^2} \\ RRMSE(u) &= \frac{\sqrt{\sum_{k=1}^{N_t} (u(\mathbf{x}) - u^*(\mathbf{x}))^2}}{\sqrt{\sum_{k=1}^{N_t} (u(\mathbf{x}))^2}} \\ E_\infty(u) &= \max_{1 \leq k \leq N_t} \|u(\mathbf{x}) - u^*(\mathbf{x})\| \end{aligned}$$

The following validation schema was implemented. First, the 8 different domains on mesh was selected for monitoring purposes. All domains for testing points were selected from the boundary conditions or at fixed distance from the borehole center. One on the inner wall of pipe r_{p-} , one on the outer wall of pipe r_{p+} , three on the borehole radius r_{b-}, r_b, r_{b+} and others are fixed distance from the borehole center $r = r_c$. Second, the time step value was selected for proper time step ($\delta\tau = 0.5$). Third, the different number of allocation points was generated on the boundary and internal domain of simulation area, in total from 100 to 1000. Forth, the distribution of allocation points could be regular, random, perturbed, showing the accuracy and quality of estimates. Fifth, one of the RBF would be selected as Multiquadratic (MQ), Inverse Multiquadratic (IMQ), Inverse Quadratic (IQ), Bessel RBF (BRBF). The testing results what RBF approximates the temperature function, also value of the parameter ϵ is should be valid. How we performed the simulation it is described in details below. The total number $N = 100, 400, 2500$ of allocation points was selected to get the optimal quality of results. The ϵ values of RBF parameter was selected from the interval $[0.1, 0.2]$. The time step is selected as the most appropriate for simulation using the fraction like $\frac{\epsilon}{\Delta} = \text{const.}$

TABLE II
THE LIST OF THERMAL PARAMETERS

Thermal conductivity, Thermal diffusivity	
$\lambda(r), \alpha(r) = \begin{cases} \lambda_{fluid}, \alpha_{fluid} & r \leq r_{p-} \\ \lambda_{pipe}, \alpha_{pipe} & r = r_p \\ \lambda_{ground}, \alpha_{ground} & r_{p+} \leq r \leq r_{b-} \\ \lambda_{soil}, \alpha_{soil} & r_{b+} \leq r \leq R \end{cases}$	

suggested by [26] and $\frac{\Delta\tau}{\Delta t} = 100$. The duration of simulation should be no less than 100 hours.

VI. CONCLUSION

Borehole heat transfer problems can be solved by the radial basis collocation methods. Results shown the shape parameter of RBF affects the accuracy of the solution. The shape parameter value of RBF is suggested. Is shown that the accuracy of solutions is more accurate then the grid spacing and time step decrease. The u_j^n solutions of selected test points gave the same order errors using the finite elements and RBF methods. Acceptance of these methods depends on more testing and application to real data. The RBF method have potential to be acceptable numerical method solving transient and steady-state borehole heat transfer problems. The further steps of investigation of RBF method would be the application to real TRT data and three-dimensional BHE modelling solving transient and steady-state borehole heat transfer problems. The simulation of BHE transfer expands the knowledge about surrounding ground and hydrogeological conditions on different layers in the three-dimensions.

APPENDIX A TABLES OF PARAMETERS

APPENDIX B NODAL DISTRIBUTION

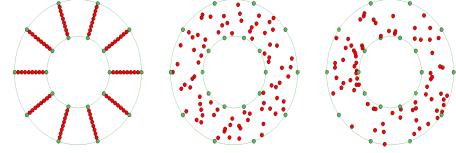
The geometrical flexibility is one of the most important advantages of the meshless methods. Such methods are expected to work well when the problem domain is represented using scattered nodes. In our numerical experiments with two- and three-dimensional examples, the interval $[r_{p-}, R]$ was represented by N^{reg} , N^{rand} , N^{per} regularly (uniformly), randomly and perturbed distribution of nodes, where N is an arbitrary positive integer. The numbers of nodes representing the domain Ω and its boundary Γ . In two-dimensional example, the boundary Γ was represented by two nodes, while in three-dimensional example the boundary Γ consisted of two concentric circles. On each circle, nodes were distributed regularly (uniformly), randomly and perturbed. The nodes representing the boundary Γ were fixed, while for the representation of the domain Ω three different strategies were used.

A. Regular distribution

In one-dimensional example, the domain Ω is represented by equidistantly distributed nodes

$$x_k = \frac{kL}{N-1}, \quad k = 1, \dots, N-2.$$

Fig. 2. Distribution of collocation points: regular, random, perturbed



In two-dimensional example, nodes are distributed regularly on concentric circles:

$$x_{kl} = (\rho_k \cos \theta_l, \rho_k \sin \theta_l),$$

where $\rho_k = r_p + \frac{k(R-r_p)}{N-1}$, $k = 1, \dots, N-2$, $\theta_l = \frac{2\pi l}{N}$, $l = 0, \dots, N-1$.

B. Perturbed distribution

Regularly distributed nodes are randomly perturbed. Therefore, the one-dimensional domain Ω is represented by the nodes

$$x_k = \frac{kL}{N-1} + \text{rand}(x_k)$$

, where $\text{rand}(x_k)$ are random numbers uniformly distributed on the interval $\left[\frac{-L}{2(N-1)}, \frac{L}{2(N-1)}\right]$, $k = 1, \dots, N-2$. In three-dimensional example (when the domain Ω is two-dimensional), regularly distributed nodes are perturbed radially and angularly:

$$x_{kl} = ((\rho_k + \text{rand}(\rho_{kl})) \cos(\theta_l + \text{rand}(\theta_{kl})), (\rho_k + \text{rand}(\rho_{kl})) \sin(\theta_l + \text{rand}(\theta_{kl}))), \quad (20)$$

where $\text{rand}(\rho_{kl})$ and $\text{rand}(\theta_{kl})$ are random numbers uniformly distributed on the intervals $\left[\frac{-(R-r_p)}{2(N-1)}, \frac{R-r_p}{2(N-1)}\right]$ and $\left[-\frac{\pi}{N}, \frac{\pi}{N}\right]$, respectively, $k, l = 0, 1, \dots, N-1$. Perturbed distribution usually is quite uniform (each node is located on its own sector).

C. Random distribution

The nodes representing the domain Ω are selected randomly. In two-dimensional example, we set $x_k = \text{rand}(x_k)$, where $\text{rand}(x_k)$ are random numbers uniformly distributed on the interval $[0, L]$, $k = 1, 2, \dots, N-2$. The two-dimensional domain Ω is represented by the following nodes:

$$x_{kl} = (\text{rand}(\rho_{kl}) \cos(\text{rand}(\theta_{kl})), \text{rand}(\rho_{kl}) \sin(\text{rand}(\theta_{kl})))$$

where $\text{rand}(\rho_{kl})$ and $\text{rand}(\theta_{kl})$ are random numbers uniformly distributed on the intervals $[r_{p-}, R]$ and $[0, 2\pi]$, respectively, $k = 1, 2, \dots, N-2$, $l = 0, 1, \dots, N-1$. If the total number of nodes is relatively small, the perturbed and random distributions of nodes can be significantly different. The density of nodes can be estimated by the fill distance which is defined as $h = \sup(x, y) \in \Omega \min_{(x_k, y_k) \in \Xi} \|(x, y)(x_k, y_k)\|^2$. The fill distance h together with the shape parameter ϵ affect the convergence rate and error estimates. Insufficiently dense distribution of the nodes can cause the increasing of the fill distance and the decreasing of the accuracy of the method. If

the domain is represented by relatively large number of nodes, the perturbed and random representations are expected to be very similar and there should not be any significant difference between the results obtained using the same method with these two strategies of the domain representation.

ACKNOWLEDGMENT

The authors would like to thank associate professor dr. Svajūnas Sajavičius from Vilnius University the faculty of mathematics and informatics for the valuable and time consuming discussions, showing the applicability of meshless algorithms.

REFERENCES

- [1] H. Carslaw and J. C. Jager, *Heat conduction in solids*. Oxford: Clarendon Press, 1st ed., 1947.
- [2] ———, *Heat conduction in solids*. Oxford: Clarendon Press, 2nd ed., 1959.
- [3] L. R. Ingersoll and H. J. Plass, “Theory of the ground pipe source for the heat pump,” *ASHRAE Trans.*, vol. 54, pp. 339–348, 1948.
- [4] J. E. Bose, “Closed-loop ground-coupled heat pump design: Manual,” *The University*, 1984.
- [5] D. P. Hart and R. Couvillion, *Earth coupled heat transfer*. Publication of National Water Well Association, 1986.
- [6] S. Kavanaugh, “Simulation and experimental verification of vertical ground-coupled heat pump systems,” Ph.D. dissertation, Oklahoma State University, Stillwater, USA, 1985.
- [7] J. Deerman and S. Kavaught, “Simulation of vertical u-tube ground coupled heat pump systems using the cylindrical heat source solution,” *ASHRAE Transactions*, vol. 97, no. 1, pp. 287–295, 1991.
- [8] G. Hellström, “Thermal analysis of duct storage system,” Ph.D. dissertation, Department of Mathematical Physics, University of Lund, Sweden.
- [9] M. A. Bernier, “Pumping design and performance modeling of geothermal heat pump systems-ground-coupled heat pump system simulation,” *ASHRAE Trans.*, vol. 107, no. 1, pp. 605–616, 2001.
- [10] M. G. Sutton, R. J. Nutter, and R. K. Davis, “An algorithm for approximating the performance of vertical bore heat exchangers installed in a stratified geological regime,” *ASHRAE Trans.*, vol. 108, no. 2, pp. 177–184, 2002.
- [11] S. Javed, “Thermal modelling and evaluation of borehole heat transfer,” Doctoral thesis, Lund University, Sweden, 2012.
- [12] Y. Man H. Yang, and J. Wang, “Study on hybrid ground-coupled heat pump system for air-conditioning in hot-weather areas like hong kong,” *Applied Energy*, vol. 87, no. 9, pp. 2826–2833, 2010.
- [13] V. C. Mei and C. J. Emerson, “New approach for analysis of ground-coil designfor applied heat pump systems,” *ASHRAE Trans.*, vol. 91, 1985.
- [14] N. K. Muraya, “Numerical modeling of the transient thermal interference of vertical u-tube heat exchanges,” 1994.
- [15] W. A. B. S. P. Rottmayer and J. W. Mitchell, “Simulation of single vertical u-tube ground heat exchanger in infinite medium,” *ASHRAE Trans.*, vol. 103, no. 2, pp. 439–445, 1997.
- [16] J. W. Thornton, T. P. McDowell, J. A. Shonder, P. J. Hughes, D. Pahud, and G. A. Hellström, “Residential vertical geothermal heat pump system models: calibration to data,” *ASHRAE Trans.*, vol. 103, pp. 660–674, 1997.
- [17] H. Zeng, N. Diao, and Z. Fang, “Heat transfer in vertical ground heat exchangers,” *International Journal of Heat and Mass Transfer*, vol. 46, no. 23, pp. 4467–4473, 2003.
- [18] R. Al-Khoury, P. G. Bonnier, and R. B. J. Brinkgreve, “Efficient finite elements formulation for geothermal heating systems.part i:steady-state,” *International journal for numerical methods in engineering*, vol. 63, no. 7, pp. 988–1013, 2005.
- [19] M. He, “Numerical modelling of geothermal borehole heat exchange systems,” Doctoral thesis, DeMontford University, 2012.
- [20] O. M. Alifanov, *Inverse Heat Conduction Problems*. Springer, 1994.
- [21] J. V. Beck, B. BlackWell, and C. R. St.Clair, *Inverse Heat Conduction Ill-Posed Problems*. John Wiley & Sons, 1985.
- [22] J. R. Cannon, *One-Dimensional Heat Equation*. Addison-Westley, 1984.
- [23] P. Hansen, *Rank-Deficient and Discrete Ill-Posed Problems*. Pa, USA: SIAM, 1998.
- [24] P. C. Hansen, “Regularization tools: a matlab package for analysis and solution of discrete ill-posed problems,” *Numerical Algorithms*, vol. 6, no. 1–2, pp. 1–35, 1994.
- [25] G. H. Golub, M. Heath, and G. Wahba, “Generalized cross validation as a method for choosing a good ridge parameter,” *Technometrics*, vol. 21, no. 2, pp. 215–223, 1979.
- [26] “Newton iteration with multiquadratics for the solution of nonlinear PDE,” *Computers & Mathematics with Applications*, vol. 43, no. 35, pp. 423–438, 2002.



Audrius Indriulionis PhD student since the year 2013 at Nature Research Center. Since the year 2002 Master Degree in Applied Mathematics the faculty of mathematics and informatics at Vilnius University. The fields of interest are dynamic partial differential equations, machine and statistical learning, time series prediction, neural networks with applications.



Juozas Bielskus Juozas Bielskus PhD student since the year 2012 at Vilnius Gediminas Technical University, Vilnius. Since the year 2012 Master Degree in energy engineering, the faculty of environmental engineering at Vilnius Gediminas Technical University. The fields of interest are applied thermodynamics, heat and mass transfer, aerodynamics.



Pranas Vaitkus Associate Professor Doctor of Mathematical Science at Department of Mathematical Statistics the faculty of mathematics and informatics at Vilnius University. The fields of interest are neural networks, machine and statistical learning, evolution algorithms.

Organizacijos modelių duomenų pakartotinis panaudojimas veiklos procesų modeliavime

Lina Valiukonytė

Informacijos sistemų katedra
Kauno technologijos universitetas
Kaunas, Lietuva
linavaliuk@gmail.com

Lina Bisikirskienė

Informacijos sistemų katedra
Kauno technologijos universitetas
Kaunas, Lietuva
lina.bisikirskiene@ktu.edu

Santrauka — straipsnyje nagrinėjamas organizaciniuose modeliuose esančios informacijos pakartotinis panaudojimas modeliuojant veiklos procesus (naudojant BPMN standartą). Pakartotinis panaudojimas realizuotas susiejant organizacinių modelių elementų tipus su veiklos procesų elementų tipais bei atliekant elementų transformaciją į naują BPMN modelį. Metodas užtikrina susiejimo schemų valdymą, leidžia pakartotinai panaudoti turimas susiejimo schemas naujiems organizacijos modeliams perkelti. Tokiu būdu sumažina analitiko darbo laiko sąnaudas bei užtikrina duomenų vientisumą kuriant skirtinges organizacijos modelius.

Raktiniai žodžiai — veiklos procesų modelis; BPMN; organizacijos modelis; susiejimo schema; transformavimas.

I. ĮVADAS

Organizacinių modelių aprašo įmonių struktūrą, procesus, žmones, kultūrą, valdymo sistemas ir technologijas. Šie modeliai šiandien yra neatsiejama verslo dalis. Informacinių sistemų analitikas modeliuoja įvairius organizacinius modelius (pvz., UPDM (angl. *Unified profile for DoDAF and MODAF*), TOGAF (angl. *The Open Group Architecture Framework*), UML (angl. *Unified Modeling Language*), SysML (angl. *Systems Modeling Language*), SBVR (angl. *Semantics of Business Vocabulary and Business Rules*) ir pan.). Šiame straipsnyje bus kalbama apie tuos organizacinius modelius, kurie yra sudaromi prieš veiklos procesų modelį, turint omeny tai, kad veiklos procesų modelis taip pat yra organizacinius modelis.

Sudarant įvairius organizacijos modelius, yra tikslinga turimą informaciją pakartotinai panaudoti sudarant naujus detalesnius modelius. Informacija yra užfiksuojama modelyje naudojant įvairius skirtinges standartus. Kadangi kiekvienas standartizuotas modelis turi savo metamodelį, taigi formuojant metamodelių susiejimo schemas atsiranda galimybė iš dalies automatizuoti būdu turimą informaciją perkelti į naujų kito standarto modelį. Perkeltą informaciją galima naudoti naujiems modeliams sudaryti. Šiuo metu dažnai informacija nėra pakartotinai panaudojama tarp skirtingu modeliavimo standartų. Taigi turimi organizacijos modeliai nėra efektyviai pritaikomi sudarant naujus modelius (pvz., veiklos procesų modeliai). Modeliavimo laikas yra ilgas, modeliai tarpusavyje nėra suderinti.

Siūlomas sprendimas leidžia turimą organizacinių modelių informaciją iš dalies automatizuotu būdu perkelti į veiklos procesų modelį. Tokiu būdu sumažinamos laiko sąnaudos modeliuojant veiklos procesus, užtikrinamas pakartotinis informacijos panaudojimas bei vientisumas.

Straipsnį sudaro keturi skyriai (organizacijos modelių analizė, organizacijos modelių duomenų pakartotinio panaudojimo veiklos procesų modeliavime metodas, išvados, literatūros sąrašas).

II. ORGANIZACIJOS MODELIŲ ANALIZĖ

Įmonės turėtų įvertinti ne tik teikiamas paslaugas, bet ir informacines sistemas, laikui bėgant jos tapo verslo dalimi. Verslo veikla kuria informacinių sistemų reikalavimus, todėl projektuoti programinę įrangą reikia pagal verslo specifiką, to nevykdžius kyla nesusipratimų pavojus. Norint suprasti kuriamos sistemos tikslą, paskirtį, reikia sukurti veiklos modelį, kurio pagrindas yra veiklos procesų aprašas. Modelis atskleidžia tikrovės pagrindinius aspektus, taip pat jame parodomos esminis tikslas, problema, kuriems skiriama tolesnis dėmesys. Teisingai suprojektuoti modeliai leidžia lengvai bendradarbiauti skirtinges kompetencijas turintiems asmenims ir kartu siekti pagrindinių tikslų.

A. Organizacijos modeliai

Standartų, skirtų organizacijai modeliuoti, yra įvairių. Vieni pagrindinių standartų yra šie: TOGAF, SysML, UPDM, UML, SBVR.

TOGAF (angl. *The Open Group Architecture Framework*) – organizacijos architektūros karkasas. Pagrindinis TOGAF komponentas – organizacijos architektūros kūrimo metodas TOGAF ADM (angl. *Architecture Development Method*). ArchiMate modeliavimo kalba yra šio karkaso pagrindas. TOGAF ADM pagrindiniai etapai yra:

- Pradinis etapas (angl. *Preliminary*),
- Reikalavimų valdymas (angl. *Requirements management*),
- A. Architektūros vizija (angl. *Architecture vision*),
- B. Veiklos architektūra (angl. *Business architecture*),
- C. IS architektūra (angl. *Information systems architectures*),

- D. Technologijų architektūra (angl. *Technology architecture*),
- E. Galimybės ir sprendimai (angl. *Opportunities and solutions*),
- F. Migracijos planavimas (angl. *Migration planning*),
- G. Igyvendinimo valdymas (angl. *Implementation governance*),
- H. Pokyčių valdymas (angl. *Architecture change management*) [7].

Veiklos procesai yra modeliuojami B. Veiklos architektūra (angl. *Business architecture*) etape.

SysML (angl. *Systems Modeling Language*) – tai yra sistemų modeliavimo kalba, kurios pagrindas yra UML modeliavimo kalba. Padidėjės sistemų sudėtingumas reikalauja daugiau sistemiškumo modeliuojant formalizuotus sistemų aprašus. Šią kalbą sudaro 9 pagrindinės diagramos, leidžiančios aprašyti sistemos struktūras bei funkcionalumą. Procesai modeliuojami veiklos diagrame (angl. *Activity diagram*) [8].

UPDM (angl. *Unified profile for DoDAF and MODAF*) – jungtinis DoDAF ir MODAF profilis. UPDM – tai organizacijos architektūros modeliavimo kalba skirta DoDAF, MODAF ir NAF architektūroms modeliuoti UML aplinkoje. UPDM standartas kuriamas OMG organizacijoje padedant JAV vyriausybei ir kitiems organizacijų nariams. UPDM sudaro 25 diagramos, iš kurių 16 skirta statinei informacijai modeliuoti, o 9 – dinaminei organizacijos veiklai. Taip pat, UPDM turi 7 aspektus (angl. *Viewpoint*):

- 1. Strateginis aspektas (angl. *Capability*),
- 2. Veiklos aspektas (angl. *Operational*),
- 3. Paslaugų aspektas (angl. *Service*),
- 4. Sistemy aspektas (angl. *Systems*),
- 5. Projektinis aspektas (angl. *Programme*),
- 6. Standartų aspektas (angl. *Technical*),
- 7. Bendrasis aspektas (angl. *All views*) [9].

Veiklos procesai yra modeliuojami veiklos aspekte (angl. *Operational*).

UML (angl. *Unified Modeling Language*) – modeliavimo ir specifikacijų kūrimo kalba, skirta specifiuoti, atvaizduoti ir konstruoti į objektą orientuotų programų dokumentus. UML yra nuosekli kalba skirta pažymeti sistemos artefaktus. Sistemos architektai gali naudoti ją apibrėžiant, vaizduojant, konstruojant ir dokumentuojant projektus. UML sudaro 15 diagramų. Veiklos procesus modeliuoti skirta veiklos diagrama (angl. *Activity diagram*) [3].

SBVR (angl. *Semantics of Business Vocabulary and Business Rules*) leidžia aprašyti organizacijos veiklos žodyną ir jos veiklos taisykles. Jis yra vystomas OMG organizacijoje. Šis standartas tiksliausiai aprašo veiklos semantiką. Jis kol kas nėra plačiai paplitęs, bet vis populiarėjantis skirtinose panaudojimo grupėse. SBVR aprašo veiklos procesams reikalingą žodyną bei veiklos taisykles [6].

Visi šie standartai suteikia galimybę kaupti organizacijos informaciją, kuri gali būti pakartotinai panaudota modeliuojant veiklos procesus. Atsižvelgiant į modelių semantiką, yra tinkamos šios standartų dalys:

- UPDM CV – strateginis aspektas, kuriame specifiuojama architektūros vizija, veiklos etapai, veiklos tikslai, veiklos funkcijos [10].
- UPDM OV – veiklos aspektas, kuriame specifiuojami veiklos objektai, veiklos procesai, veiklos aprivojimai, veiklos duomenys ir duomenų srautai [10].
- TOGAF A etape yra architektūros vizijos parengimas, darbo planas, veiklos tikslai, veiklos funkcijų įvertinimas [4].
- TOGAF B etape yra pakoreguota vizija, reikalavimų architektūrai specifikacija, architektūros aprašas, architektūros vystymo planas, organizacijų katalogai, veiklos sąveikų matricos, veiklos, panaudos atvejų, organizacinių struktūros, procesų ir įvykių diagrame [4].
- SBVR veiklos žodynas turi veiklos esybių, jų egzempliorių bei ryšių tarp jų rinkinį, kurį gali naudoti bet kokia organizacija [6].
- UML klasų, veiklos, panaudos atvejų diagrame – klasų diagrama apibūdina sistemos struktūrą parodydama sistemos klasės, jų atributus bei sąryšius tarp tų klasių. Veiklos diagrama žingsnis po žingsnio atvaizduoja sistemos ir komponentų veiksmus. Panaudos atvejų diagrama naudojama sistemą sudarantiems pirminiams elementams ir procesams nustatyti [3].

B. Veiklos procesas ir jo modeliavimo standartai

Veiklos procesai yra sudedamoji kiekvieno iš anksčiau apibrėžtų standartų dalis. Nepaisant to, kad jiems modeliuoti yra naudojama skirtinga notacija, tačiau pagrindiniai principai išlieka tie patys.

Pagal G. A. Rummlerį ir A. P. Brache, veiklos procesas yra serija žingsnių, suprojektuotų gaminti produktą ar paslaugą. Pagal OMG, veiklos procesas yra aibė apibrėžtų veiklų, kurios reprezentuoja reikalingus žingsnius, kad būtų pasiekta verslo tikslas, rinkinys [1].

Kaip minėta anksčiau, veiklos procesai yra modeliuojami šiuose standartuose: UML, SysML, UPDM, TOGAF. Apibendrinus gerąją praktiką, buvo sudarytas atskiras standartas jų modeliavimui, t. y. BPMN. Šiuo metu, jo diagramos dažnai integruojamos minėtuose standartuose (pvz., UPDM, TOGAF).

BPMN (angl. *Business Process Model and Notation*) – tai OMG standartas, skirtas analizuoti organizacijos veiklai taikant grafinę notaciją, komunikuoti, aprašyti verslo sandorius (angl. *Business transactions*) tarp organizacijų. Pagrindinės BPMN elementų grupės:

- valdymo srauto objektai (angl. *Flow objects*),
- jungiantys objektai (angl. *Connecting objects*),
- sritys (angl. *Swimlanes*),
- artefaktai (angl. *Artifacts*),
- duomenys (angl. *Data*) [2].

Pagal BPMN standartą parengti modeliai leidžia organizacijoms suprasti savo vidaus veiklos procedūras grafinėje notacijoje. Grafinė notacija palengvina darbo bendradarbiavimo supratimą. BPMN leidžia suprasti įmonėms

jų veiklos procedūras, dalyvius dalyvaujančius jų procesuose ir prisitaikyti naujam veiklos modeliavimui. Taip pat, nuolatinis gerinimas, orientuotas į rezultatą ir klientą. Procesai skirstomi į pagrindinius ir palaikančiuosius. Leidžia automatizuoti procesus panaudojant informacines sistemas.

Išanalizavus pagrindines veiklos procesų modeliavimo galimybes, buvo nuspėsta pasirinkti BPMN veiklos procesų modeliavimo kalbą kaip tikslinę modeliavimo kalbą į kurią transformuojama kituose modeliuose turima informacija.

III. ORGANIZACIJOS MODELIŲ DUOMENŲ PAKARTOTINIO PANAUDOMIMO VEIKLOS PROCESŲ MODELIAVIME METODAS

Šis metodas skirtas organizacijos modelių duomenų (elementų) pakartotiniams panaudojimui modeliuojant veiklos procesus. Metodas susideda iš dviejų pagrindinių etapų:

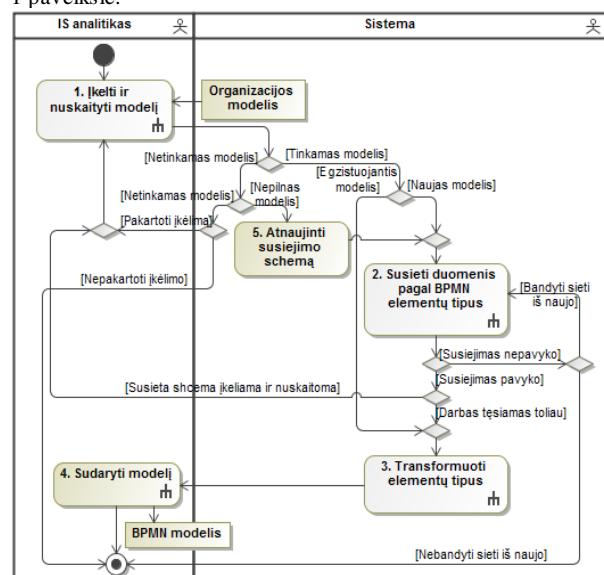
- pirmio modelio elementų tipų susiejimas su BPMN elementų tipais. Sudaroma susiejimo schema. Susiejimas yra loginis, t.y. turi atitinkti elementų tipų semantika.
- elementų transformavimas pagal taisykles ir BPMN pradinio modelio sudarymas.

Rezultatas gali būti plačiai naudojamas, kadangi susiejimus susiejimo schemaje ekspertas gali atlikti be apribojimų, t. y. be apribojimų naudoti turimą organizacijos modelį. Taip informacija yra panaudojama efektyviau. Šiam tikslui pasiekti vartotojas pirmiausiai turi turėti sukurtą organizacijos modelį (pvz., UPDM modelis), iš kurio išgauti elementų tipai yra susiejami su BPMN elementų tipais bei įvykdoma transformacija.

Metodas detaliau aprašytas tolimesniuose poskyriuose.

A. Algoritmas

Pagrindinė algoritmo veiklos diagrama yra pateikiamai 1 paveiksle.

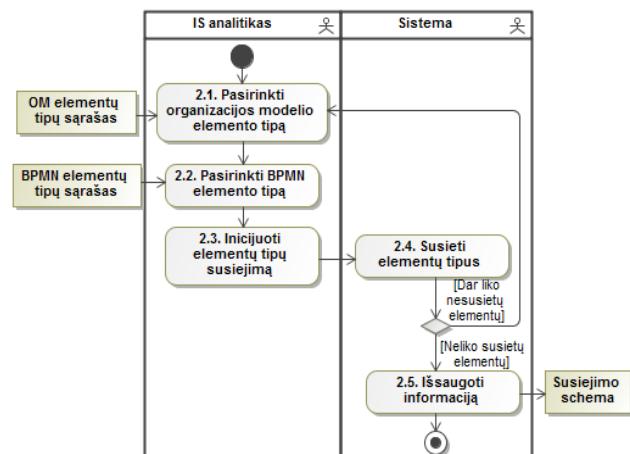


Pagrindiniai algoritmo žingsniai yra šie:

- 1) *ikelti ir nuskaityti modelį;*
- 2) *susieti duomenis pagal BPMN elementų tipus;*
- 3) *transformuoti elementų tipus;*
- 4) *sudaryti modelį;*
- 5) *atnaujinti susiejimo schemą.*

1 žingsnis. Ikelti ir nuskaityti modelį

Jau turimą organizacijos modelį reikia eksportuoti į XMI failą. Failas įkeliamas į sistemą ir jeigu yra teisingo formato, nuskaitomi elementų tipai.



2 žingsnis. Susieti duomenis pagal BPMN elementų tipus

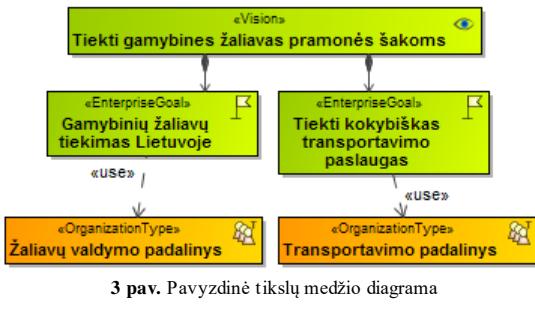
Tam, kad susietume elementų tipus susiejimo schemaje, turime turėti nuskaitytą organizacijos modelio elementų tipų sąrašą ir sąrašą BPMN elementų tipų (šis sąrašas nėra ribojamas, t. y. įtraukti visi galimi BPMN elementų tipai). IS analitikas pagal dalykinę sritį ir pagal savo kompetenciją susieja elementų tipus, t. y. iš kokių į kokius nori atlikti transformaciją (2 pav.).

Kai susiejimas baigtas informacija išsaugoma ir turime išsaugotą susiejimo schemą, kurią ateityje bus galima panaudoti arba atnaujinti. Susiejimo schemas galima naudoti ir kitiems to pačio tipo modeliams transformuoti.

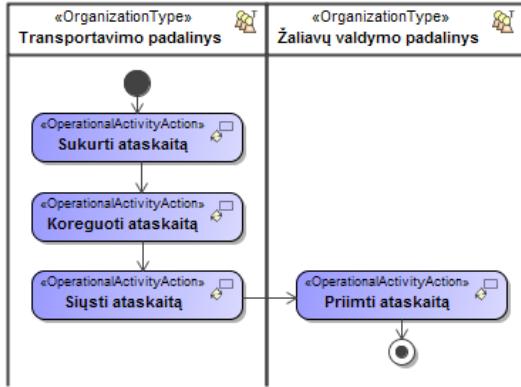
3 žingsnis. Transformuoti elementų tipus

Kai jau turime susiejimo schemą, reikia įvykdyti elementų tipų transformavimą. Transformavimo rezultatai yra rodomi sistemoje. Jis atliekamas vieną elemento tipą pakeičiant kitu ir paliekant tą patį elemento pavadinimą. Rezultatai pateikiami XMI formatu.

Pavyzdiniame projekte naudojamos UPDM standarto diagramos. Modelyje buvo sukurtos tikslų medžio (3 pav.) ir veiklos diogramos (4 pav.).



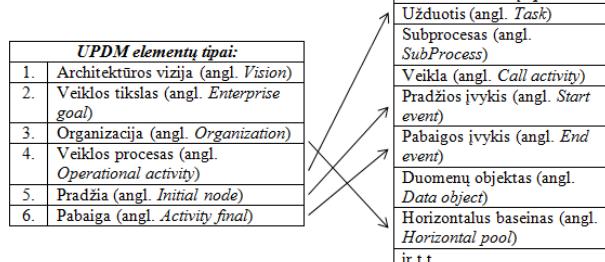
3 pav. Pavyzdinė tikslų medžio diagrama



4 pav. Pavyzdinė veiklos diagrama

Projektas eksportuotas į XMI formatą. IS analitikui yra suteikiama galimybė susieti elementų tipus be aprūpoinimų. Kadangi jis yra ekspertas, elementų tipus gali susieti pagal modeliuojamos srities kontekstą. Tokiu būdų net ir tam pačiam standartui galima sukurti kelias susiejimo schemas.

Šiame pavyzdyste iš UPDM standarto modelio elementų tipų pasirinkti „organizacija“, „veiklos procesas“, „pradžia“, „pabaiga“, kurie transformuojami į BPMN standarto „užduoties“, „horizontalaus baseino“, „pradžios įvykio“ ir „pabaigos įvykio“ tipo elementus.



5 pav. UPDM ir BPMN elementų tipų susiejimo pavyzdys

Elementų tipų susiejimas yra pavaizduotas 5 paveiksle. 6 paveiksle pateikiamas elementų transformavimas pagal susiejimo schema.

UPDM elementų tipai ir elementai:
1. Architektūros vizija (angl. Vision)
a. Tiekti gamybines žaliavas pramonės šakoms
2. Veiklos tikslas (angl. Enterprise goal)
a. Gamybinų žaliavų tiekimas Lietuvoje
b. Tiekti kokybiškas transportavimo paslaugas
3. Organizacija (angl. Organization)
a. Transportavimo padalinys
b. Žaliavų valdymo padalinys
4. Veiklos procesas (angl. Operational activity)
a. Sukurti ataskaitą
b. Koreguoti ataskaitą
c. Siųsti ataskaitą
d. Priimti ataskaitą
5. Pradžia (angl. Initial node)
6. Pabaiga (angl. Activity final)

Transformuoti elementai į BPMN:

- | |
|--|
| 1. Horizontalus baseinas (angl. Horizontal Pool) |
| a. Transportavimo padalinys |
| b. Žaliavų valdymo padalinys |
| 2. Užduotis (angl. Task) |
| a. Sukurti ataskaitą |
| b. Koreguoti ataskaitą |
| c. Siųsti ataskaitą |
| d. Priimti ataskaitą |
| 3. Pradžios įvykis (angl. Start event) |
| 4. Pabaigos įvykis (angl. End event) |

6 pav. UPDM elementų tipų transformavimas į BPMN elementų tipus

4 žingsnis. Sudaryti modelį

Kuomet jau atliktas transformavimas, turime sugeneruotą failą su perkeltais elementais. Tuomet jau galima panaudoti BPMN elementus iš elementų bibliotekos (XMI). Failas yra įkeliamas į MagicDraw programą.

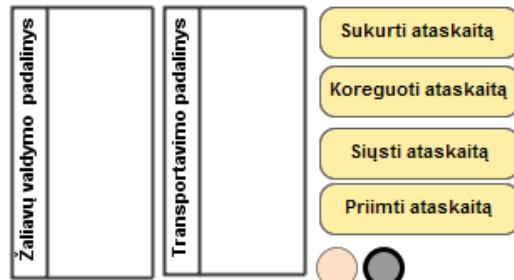
Visa išsaugota ir transformuota informacija vartotojui pateikiama lentelėje (7 pav.), kurioje matomi elementų tipai, elementų pavadinimai ir jų būsena (ar elementas buvo susietas ir į koki elemento tipą buvo transformuota). Prototipas realizuotas C# programavimo kalba.

Elementų transformavimas		
OM elemento tipas	OM elemento pavadinimas	OM elemento transformavimo būsena
Vision	Tiekti gamybines žaliavas pramon...	Nesusiepta
Enterprise goal	Gamybinų žaliavų tiekimas Lietuv...	Nesusiepta
Enterprise goal	Tiekti kokybiškas transportavimo p...	Nesusiepta
Organization	Transportavimo padalinys	Transformuota į tipą "Horizontal Pool"
Organization	Žaliavų valdymo padalinys	Transformuota į tipą "Horizontal Pool"
Operational activity	Sukurti ataskaitą	Transformuota į tipą "Task"
Operational activity	Koreguoti ataskaitą	Transformuota į tipą "Task"
Operational activity	Siųsti ataskaitą	Transformuota į tipą "Task"
Operational activity	Priimti ataskaitą	Transformuota į tipą "Task"
Initial node		Transformuota į tipą "Start Event"
Activity final		Transformuota į tipą "End Event"

Pradėti transformavimą

7 pav. Elementų tipų susiejimo rezultatai

8 paveiksle demonstruojami transformuoti ir įkelti į BPMN diagramą elementai.



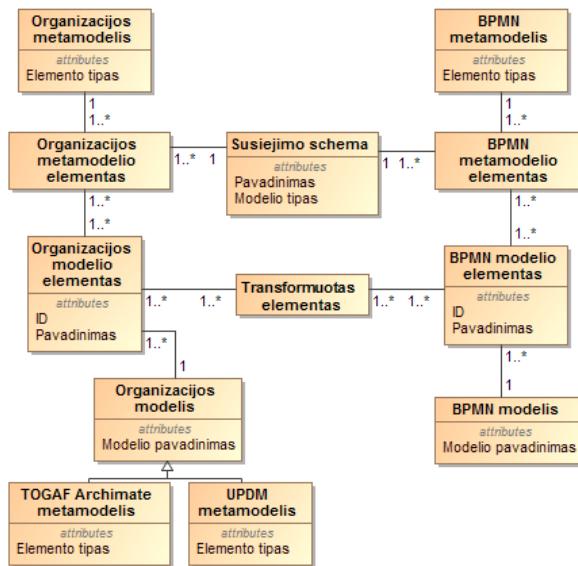
8 pav. Transformuoti elementai BPMN diagramoje

5 žingsnis. Atnaujinti susiejimo schema

Iškėlus organizacijos modelį į irankį, yra pasirenkamas modelio tipas. Irankis turi galimybę valdyti susiejimo schemas pagal modelio tipus. Kuomet elementų tipai yra susiejami, susiejimo schema yra išsaugoma su nurodytu modelio tipu. Tokiu būdu yra galimybė panaudoti susiejimo schemą ikeliant naujus organizacijos modelius, ją atnaujinti. Susiejimo schemų gali būti ne viena, jos skirstomos pagal modelio tipus. Atnaujinus arba panaudojus susiejimo schemą elementai yra transformuojami.

B. Dalykinės srities esybių klasinių modelis

Algoritmo realizacijai sudarytas dalykinės srities esybių klasinių modelis (9 pav.) atvaizduoja kuriamo algoritmo pagrindinius konceptus (Organizacijos metamodelis, organizacijos metamodelio elementas, susiejimo schema, BPMN metamodelis, BPMN metamodelio elementas, organizacijos modelio elementas, transformuotas elementas, BPMN modelio elementas, BPMN modelis, organizacijos modelis, TOGAF ArchiMate metamodelis, UPDM metamodelis) ir duomenis.



9 pav. Dalykinės srities esybių modelis

IV. IŠVADOS

Pakartotinis panaudojimas realizuotas susiejant organizacinių modelių elementų tipus su veiklos procesų elementų tipais bei transformuojant elementus į BPMN modelį. Metodas užtikrina susiejimo schemų valdymą, leidžia pakartotinai panaudoti turimas susiejimo schemas naujiems organizacijos modeliams, tokiu būdu sumažina analitiko darbo laiko sąnaudas bei užtikrina duomenų vientisumą kuriant skirtinges organizacijos modelius. IS analitikas nebūtinai turi naudoti BPMN notaciją, galima rinktis be apribojimų. Esant poreikiui galima atnaujinti susiejimo schemas. Irankis skirtas IS analitikui ekspertui, nes reikalauja kvalifikacijos, kad susiejimai būtų prasmingi. Gautos duomenų pakartotino panaudojimo algoritmas gali būti sėkmingesni panaudotas tolimesniuose IS kūrimo etapuose ir organizacijos veikloje.

V. LITERATŪRA

- [1] T. Weilkiens, C. Weiss, A. Grass, „OCEB Certification Guide“ Business Process Management – Fundamental Level, 2011.
- [2] Emerald group publishing, „Business process modelling, simulation and reengineering: call centres“, 2008. [Tinkle]. Available: <http://www.emeraldinsight.com/doi/pdfplus/10.1108/14637150810916017>. [Žiūrėta 2014 10 17].
- [3] Unified Modeling Language® (UML®) Resource Page. [Tinkle]. Available: <http://www.uml.org/> UML [Žiūrėta 2016 02 25].
- [4] ArchiMate® 2.1 Specification, 2013. [Tinkle]. Available: <http://pubs.opengroup.org/architecture/archimate2-doc/chap08.html>. [Žiūrėta 2015 02 28].
- [5] Duomenų projektavimas, VGTU. 2010. [Tinkle]. Available: http://www.techmat.vgtu.lt/konspektaiPSI/psi_09_duomenų_projektavimas.pdf [Žiūrėta 2016 02 25].
- [6] E. Šinkevičius, L. Tutkutė, SBVR metamodelio sudarymas ir panaudojimas veiklos žodynyje transformacijoms į UML taikant ATLAS transformavimo kalbą, mokslinis straipsnis, IVUS 2010 medžiaga. 2010 m. gegužės 13 d., Kaunas.
- [7] TOGAF. [Tinkle]. Available: <http://pubs.opengroup.org/architecture/togaf9-doc/arch/toe2.html> [Žiūrėta 2015 01 27].
- [8] Friedenthal, Moore, & Steiner, OMG Systems Modeling Language, 2008. [Tinkle]. Available: <http://www.omg.org/INCOSE-OMGSysML-Tutorial-Final-090901.pdf> [Žiūrėta 2015 01 25].
- [9] Cameo enterprise architekture user guide, MoMagic, Inc. 2010. [Tinkle]. Available: http://www.magicdraw.com/files/manuals/beta/Cameo%20EA%20User_Guide.pdf [Žiūrėta 2016 03 01].
- [10] Unified Profile for DoDAF and MODAF (UPDM), OMG, 2013. [Tinkle]. Available: <http://www.omg.org/spec/UPDM/2.1/PDF/> [Žiūrėta 2016 02 25].

Mašinų, teikiančių informaciją *MTConnect* protokolu, duomenų surinkimas ir analizė

Ignas Sedunovas

Programų inžinerijos katedra

Kauno technologijos universitetas

Kaunas, Lietuva

ignas.sedunovas@ktu.edu

Santrauka – šiame straipsnyje yra aprašomas *MTConnect* standarto panaudojimas renkant pramoninių staklių duomenis ir juos analizuojant, siekiant identifikuoti artėjantį gedimą. Pateikiama keletas jau esamų sprendimų ir siūlomas sprendimas, pristatomos problemos, iškilusios jį įgyvendinant. Taip pat aprašomas planuojanamas tyrimas, kaip pagerinti siūlomo sprendimo duomenų analizės greitaveiklą manipuliuojant, konvertuojant duomenis ir optimizuojant koreliacių skaičiavimus.

Raktiniai žodžiai – *MTConnect*; pramoninės staklės; didieji duomenys (angl. big data); duomenų analizė.

I. ĮVADAS

Šiais laikais yra stengiamasi įvairius procesus automatizuoti. Gamybos pramonėje yra pasitelkiamos mašinos / robotai, kurie gamina produkta pagal tam tikrą kompiuterinę programą. Kurį laiką mašinų komunikacijai buvo rašomos specifinės programos, kurių reikėdavo nemažai ir jas būdavo nelengva palaikti. Kad tai palengvintų, buvo standartizuotas nemokamas mašinų duomenų perdavimo internetu protokolas *MTConnect* [1], kurio pirmoji versija 1.0.1 išleista 2009 metais [2].

Nuo to laiko, kai atsirado standartizuotas metodas gauti mašinų duomenis nuotoliniu būdu, mašinų pramonėje iškilo ir vis dar kyla poreikis per atstumą stebeti ir analizuoti duomenis, kuriuos pateikia veikiantys įrengimai. To įmonėms reikia dėl to, kad sumažėja laiko, darbo jėgos ir išlaidų sąnaudos, didėja produktyvumas, atsiranda geresnės sąlygos konkuruoti, ir galima bet kuriuo metu prižiūrėti mašinas iš bet kurios vienos su interneto prieiga [3] [4].

Vienas iš straipsnio tikslų yra aprašyti sprendimą, kurį naudojant būtų surenkami pramoninių staklių rodmenys ir analizuojami siekiant nustatyti pasikartojančius gedimus prieš jiems dar įvykstant. Kitas tikslas – aprašyti tyrimo metodus, kuriais siekiama manipuliuoti, konvertuoti duomenis ir optimizuoti koreliacių skaičiavimus.

Šis straipsnis susideda iš supažindinimo su *MTConnect* standarto veikimo principu, įvairių pramonei skirtų sprendimų, naudojančių *MTConnect*, apžvalgos, sukurto sprendimo ir problemų įgyvendinant jį aprašymo. Taip pat yra pateikiamas planuojanamas tyrimas sukurtos sistemos našumui gerinti bei išvados.

II. MTCONNECT STANDARTAS

MTConnect yra atviras, nemokamas standartas, skirtas skatinti didesnę sąveiką tarp įrenginių ir programinės įrangos. Jis yra pagrįstas atviruoju duomenų integracijos protokolu. Sukurdamas atvirą ir išplečiamą komunikacijos kanalą savaiminio diegimo (angl. plug-and-play) tarpusavio ryšiui tarp prietaisių, įrangos ir sistemų, *MTConnect* leidžia jiems keistis ir suprasti vienį kitų duomenis ir šitaip sumažinti integracijos kaštus. Akcentuotina, kad standartas skirtas perduoti duomenis ir kad jis duomenims nesuteikia jokios papildomos prasmės.

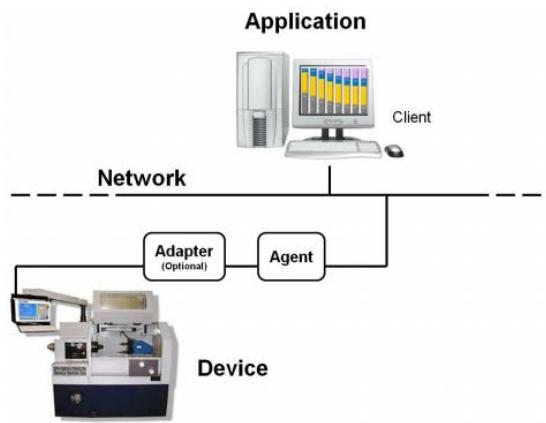
MTConnect yra sukurtas remiantis labiausiai paplitusiais pramonės ir programinės įrangos standartais, todėl maksimaliai padidina jam įgyvendinti prieinamų įrankių kiekį ir suteikia aukščiausią sąveiką su kitais pramonės šakų standartais ir įrankiais. Pagrindiniai technologiniai komponentai, naudojami *MTConnect*, yra universaliojo dokumentų ženklinimo kalba (*XML*), skirta dokumentų struktūrai aprašyti, ir hipertekstų persiuntimo protokolas saityno duomenims (ištekliams) persiųsti (*HTTP*) [5].

Sistema, įgyvendinta naudojant *MTConnect* standartą, yra sudaryta iš 5 pagrindinių dalių (1 pav.) [5]:

- **Įrenginys** (angl. *device*) – dažniausiai mašininis įrankis, bet gali būti bet kuri įrangos dalis arba duomenų šaltinis.
- **Adapteris** (angl. *adapter*) – neprivaloma programinės arba techninės įrangos dalis, kuri konvertuoja rodmenis į *MTConnect* standartą. Kai kurie įrenginiai rodmenis teikia jau pagal standartą – tokiu atveju adapteris nereikalingas.
- **Agentė** (angl. *agent*) – programinės įrangos dalis, kuri surenka, organizuoja ir kaupia duomenis, gautus iš įrenginio ar adapterio. Ji apdoroja užklausas, gautas iš vartotojo (taikomosios programos), ir pateikia reikiamus duomenis. Agentės funkcija ir duomenų struktūra yra specifikuota *MTConnect* standarto.
- **Tinklas** (angl. *network*) – fizinis ryšys tarp duomenų šaltinio (įrenginio) ir duomenų naudotojo (taikomosios programos). Dažniausiai tai yra interneto tinklas ir išprastai yra naudojamas standartinis bendravimo metodas – *HTTP* protokolas. Be to, *MTConnect*

struktūra gali būti pritaikyta ir kitokiems tinkliniams sprendimams.

- Taikomoji programa (angl. *application*) – tikrasis *MTConnect* duomenų prašytojas ir naudotojas. Tipinės taikomosios programos funkcijos yra teikti užklausas, kaupti, manipuliuoti ir atvaizduoti duomenis.



1 pav. *MTConnect* komponentų schema [5]

Šis standartas yra nuolat pildomas ir tobulinamas naudojant kitus technikos standartus – 2015 metų birželį buvo išleista naujausia 1.3.1 versijos specifikacija [2].

III. ESAMI SPRENDIMAI, NAUDΟJANTYS *MTCONNECT*

Nuo 2013 metų JAV gynybos departamentas rengia kasmet vykstantį *MTConnect Challenge* konkursą, skirtą skatinti kurti ir igyvendinti idėjas, kaip galima panaudoti gamybos duomenis [6]. Be to, kasmetinėje *ASME (The American Society of Mechanical Engineers)* konferencijoje taip pat yra pristatomi įvairūs sprendimai, naudojantys *MTConnect*.

A. „Microsoft Visio“ integracija [7]

Jamesas Finnas sukūrė *MTConnect* sprendimą *Microsoft Visio* programai. Šis modeliavimo įrankis leidžia gamybos kontrolės priežiūros personalui greitai ir neprikalusomai susikurti duomenų gavimo ir atvaizdavimo / peržiūros langą, kuriuo naudodamies vėliau gali išspręsti netikėtai atsiradusias gamybos cecho problemas.

Šio sprendimo pagrindinė savybė yra tai, kad naudojant *Microsoft Visio* duomenys iš viso cecho, kurie tuo metu yra transliuojami *MTConnect* kanalu, yra greitai surenkti, agreguojami ir vaizdžiai pateikiama realiu laiku.

B. Virpėjimo nustatymas panaudojant stebėseną [8]

Autorius nagrinėja anomaliją, kai mašinos pjaunamosios ar gržiamosios detalės pradėdavo virpęti per daug ir dėl to detalė nusidėvėdavo greičiau, o produkto kokybė suprastėdavo. Pažeistų detalų keitimasis kainuoja daugiau ir ilgas mašinos neveiksnumo laikas remonto metu gali neigiamai paveikti visą cecho produktyvumą. Valerie Pezzullo reikėjo būdo tai pastebėti anksčiau negu atsirasdavo defekto garsas.

Autorius sprendimas buvo panaudoti *MTConnect* duomenis apie detalių parametrus iš prietaiso ir duomenis apie vibracijos lygi iš jutiklių juos perduodant į kontrolinį valdiklį, kuris ivertinės esamą situaciją atlieka tam tikras prietaisų korekcijas (pvz., sumažina grąžto sukimosi greitį). Atlikus skaičiavimus vienam procesui, apdoroti duomenys yra siunčiami į kitą valdiklį, kuris atsakingas už kitą procesą, kad jis taip pat ivertintų šią padėtį ir atlktų reikiamus pakeitimus.

C. „Google Glass“ sprendimas [9]

Ravis Sivalingamas savo sprendime apjungė *MTConnect* funkcionalumą ir *Google Glass* galimybes: ižvalgų ekraną, vaizdo kameral, lietimui jautrią dalį, mikrofoną, el. paštą ir interneto prieigą. Nuskenavus su vaizdo kamera unikalų mašinos kodą darbuotojas akių ekrane mato to prietaiso dabartinius parametrus, gali apžiūrėti detales trimatėje erdvėje, gali modifikuoti prietaiso darbo planą ir perduoti bet kokią informaciją kolegai.

D. „Promise“

MTConnect Challenge 2 konkurse Shane’as Crandallas pristatė savo intuityvią, lengvai valdomą internetinę programą, kuri naudoja *MTConnect* teikiamą informaciją. Sprendimas vartotojui suteikia galimybę tuščiame lange susikurti įvairių matavimo formų (stulpelinė diagrama, grafikas, matuoklis su ciferblatu ir pan.) elementus ir nustatyti, kurio prietaiso ir kurie duomenys bus rodomi. Ši programa taip pat leidžia nustatyti duomenų vaizdavimo spalvą pagal reikšmę, keisti foną, kad būtų lengviau susigaudyti, kurioje prietaiso vietoje stovi matuoklis. Visi duomenys yra vaizduojami realiu laiku.

E. „Kaizen“ pobūdžio sprendimas [10]

Šis sprendimas buvo konkrečiai taikomas optimizuoti *Boeing* įmonės procesus. *MTConnect* pagalba duomenys buvo žymiai greičiau surenkti negu naudojant firminius tū mašinų įrankius. Surinkti duomenys buvo agreguojami ir pateikiami įvairių grafikų pavidalu. Juos išanalizavęs cechą aptarnaujantis personalas galėjo lengviau ir greičiau identifikuoti realią įvairių procesų trukmę, kainą ir kitus parametrus.

F. Įrenginių sąveikos gerinimas [11]

Athulanas Vijayaraghavanas kartu su komanda bandė sujungti nesuderinamus prietaisus. Naudodami *MTConnect* standartizuotą sąsają jie apjungė įvairius įrenginius bendram tikslui. Taip pat buvo renkami duomenys, kuriuos programinė įranga realiu laiku analizuodavo ir sek davavo, ar mašina veikia taip, kaip ji buvo užprogramuota (buvo vykdomas proceso plano verifikavimas). Vėliau tuos duomenis ir išvestinius rezultatus galėjo peržiūrėti cecho personalas ir atlkti tam tikrus proceso pakeitimus.

G. Komercinių ribotų SPC metodų gerinimas [12]

Statistinės procesų kontrolės (angl. *Statistical Process Control (SPC)*) tikslas yra ne aptiktis blogą kokybę, o užkirsti kelią jos atsiradimui. *SPC* metodais yra siekiama stebeti procesą, jį kontroliuoti ir atlkti reikiamus pakeitimus. Tuo tikslu autoriai naudodami *MTConnect* sukūrė programą, kuri analizuodavo gautus duomenis ir bandydavo apskaičiuoti tolimesnę prietaiso eiga. Jei buvo nustatomi galimi

nukrypimai, tai programa įvertindavo, kaip reikia pamodifikuoti tolimesnę proceso eiga ir nusiųsdavo šią komandą mašinai.

H. Gamyklos sistemų kokybės stebėjimas realiu laiku [13]

Siekdami pagerinti gamybos kokybę, kuri buvo įvertinama pasibaigus procesui, kai galutinis produktas yra apžiūrimas ir įvertinamas, o po to retrospektivai įvertinamas pats procesas ir sąnaudos, autoriai C++ kalba sukūrė taikomąją programą, kuri panaudodama MTConnect gaudavo duomenis iš CMM, juos apdorodavo, suformuodavo HTML dokumentą ir realiu laiku pateikdavo interneto naršykėje (konkrečiai, Internet Explorer). Pateiktame dokumente buvo išskiriamos reikšmės, kurios peržengdavo viršutinę arba apatinę maksimalaus leistino nuokrypį ribą. Automatizavus stebėjimo procesą, aptarnaujančiam personalui liko daugiau laiko kitoms produktyvumo užduotims.

I. „Pneuviz“ – suspausto oro stebėjimo programa [14]

Gamybos pramonėje suspaustas oras yra laikomas ketvirtą didžiausia pagalbine priemone po elektros, gamtinų dujų ir vandens. Jis yra naudojamas daugybėje įvairių pneumatiniu ir mechaninių prietaisų kiekvienam cechu. Todėl Sri Atluru su trimis kolegomis sprendė suspausto oro stebėjimo ir utilizavimo optimizavimo problemą ir sukūrė projektą, kuris išanalizuoją suspausto oro panaudojimą įvairiose situacijose metalo pjovimo operacijų metu. Taikomoji programa buvo sukurta naudojant LabVIEW programinį paketą ir skirta stebeti proceso bei analizuoti informaciją. Duomenims gauti programa buvo susieta su MTConnect duomenimis, perduodamais įmonės vidiniu tinklu. Specifinių jutiklių rodmenys buvo nagrinėjami įvairiuose scenarijuose ir taip buvo nustatoma kainos, energijos ir anglies pėdsakų įtaka suspausto oro panaudojamumui cechė.

J. „MTConnect“ pritaikymas paskirstytose gamybos aplinkoje [15]

Johnas L. Michaloskis su kolegomis sprendė problemą, kad reikia patogaus detales gaminančių programuojamų mašinų duomenų stebėjimo realiu laiku. Valdyba teigė, kad informatyvūs, tikslūs ir laiku pateikiami įrenginių duomenys gali būti kritinis sėkmingsos gamybos aspektas. Kadangi įmonės tinklas turėjo saugumo apribojimų, buvo nuspresta naudoti dviejų linijų intraneto ryšį. Pirmaja linija neapdoroti duomenys buvo perduodami iš prietaiso į atskirą kompiuterį, kuriame MTConnect pagalba jie būdavo konvertuojami į standartizuotą formatą. Iš šio kompiuterio kita linija duomenys būdavo perduodami į centralizuotą prietaisų stebėjimo skydą. Šis sprendimas įmonei garantavo tikslų gamybos kokybės įvertinimą realiu laiku.

K. Įrenginių stebėjimo sistema, grįsta „MTConnect“ technologija [16]

Sprendimas susideda iš dviejų tinklu sujungtų komponentų: paties įrenginio ir pranešimų serverio. Mašinoje esantis MTConnect adapteris surenka dabartinės būsenos duomenis ir intranetu siunčia į pranešimų serveryje esančią MTConnect agentę, kuri apdoroja gautus duomenis, konvertuoja į standartizuotą MTConnect XML formatą. Tada šiuos duomenis duomenų surinkimo servisas išsaugo į serverio duomenų bazę. Vėliau iš šių duomenų yra generuojamos ataskaitos, kurias naudojant interneto naršykę galima pasiekti prisijungus prie pranešimų serveryje esančios saityno programos. Taip pat, atsižvelgiant į surinktus duomenis, el. pašto serverio pagalba mašinų operatoriams yra siunčiami pranešimai SMS ir / arba el. laišku.

1 LENTELĖ. MTCONNECT SPRENDIMŲ PALYGINIMAS

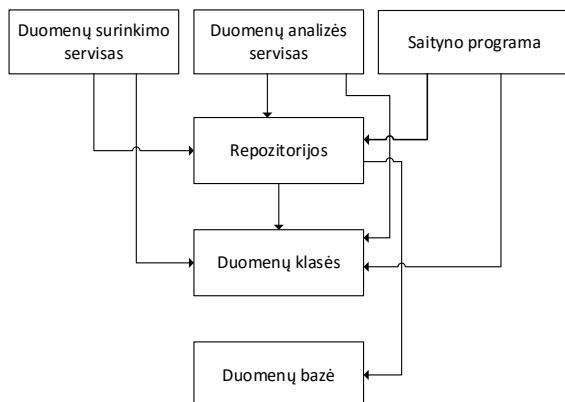
	Priima standartizuotus XML duomenis	Kaupia gautus duomenis	Duomenys yra vaizdujai pateikiamai	Duomenys pateikiamai realiu laiku	Yra galimybė įtinkinti duomenų atvaizdavimą	Analizuojama duomenis	Siunčia pranešimus	Modifikuoja proceso eigoje
„Microsoft Visio“ integracija	+	+	+	+	+	-	-	-
Virpėjimo nustatymas panaudojant stebėseną	+	-	+	+	-	+	-	+
„Google Glass“ sprendimas	+	-	+	+	-	-	+	-
„Promise“	+	+	+	+	+	-	-	-
„Kaizer“ pobūdžio sprendimas	+	+	+	+	+	-	-	-
Įrenginių saveikos gerinimas	+	+	+	+	+	-	-	-
Komerciniu ribotu SPC metodų gerinimas	+	+	+	+	-	+	-	+
Gamyklos sistemų kokybės stebėjimas realiu laiku	+	-	+	+	-	-	-	-
„Pneuviz“ – suspausto oro stebėjimo programa	+	+	+	+	+	-	-	-
„MTConnect“ pritaikymas paskirstytose gamybos aplinkoje	+	+	+	+	+	-	-	-
Įrenginių stebėjimo sistema, grįsta „MTConnect“ technologija	+	+	+	+	+	-	+	-

IV. SIŪLOMAS SPRENDIMAS

A. Igyvendinta sistema

Buvo sukurtas sprendimas rinkti ir kaupti pramoninių staklių rodmenis, kurie po to yra atvaizduojami aptarnaujančiam personalui prisijungus prie saityno programos. Siekiant atskirti skirtingo funkcionalumo procesus sistemą (2 pav.) sudaro trys dalykai (duomenų surinkimo ir analizės servisių bei saityno programa), kurių kiekviena suprojektuota pagal trijų sluoksnį architektūrą ir naudoja tą pačią duomenų bazę. Jos igyvendinimui buvo panaudotos šios technologijos: *C#.NET MVC, Elastic, NEST, Autofac, Math.NET*.

Duomenų surinkimui buvo parašytas atskiras modulis (duomenų surinkimo servisas), kurio funkcionalumas visiškai atitiko nurodymus ir reikalavimus *MTConnect* standarto specifikacijoje [1], kurių reikia norint kaupti rodmenis esant normaliomis sąlygomis ir klaidų atvejais.



2 pav. Sukurtos sistemos architektūra

Duomenų analizei sukurtas modulis kurios nepertraukiama analizuoja naujausius įrenginių rodmenis. Kadangi pagal standartą irangos gedimai yra diskrečios būsenos, kurios gaunamos kartu su irangos rodmenimis [1], todėl analizė buvo vykdoma lyginant n naujausių įrenginių rodmenų poaibį $\{x_1, \dots, x_n\}$ su istorinių, buvusių prieš gedimus rodmenų poaibiu $\{y_1, \dots, y_n\}$ ir ieškant reikšmingos koreliacijos. Rodmenų panašumas yra laikomas reikšmingu, jei Pirsono koreliacijos koeficientas r (1) gaunamas ne mažesnis už 0.7:

$$r = r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (1)$$

Viena analizės iteracija apima vieno įrenginio naujausiu rodmenų lyginimą su tokios pačios imties istoriniais rodmenimis, kurie buvo gauti prieš visus to prietaiso gedimus. Analizės metu yra ieškoma ir reikšmingų duomenų panašumų ir anomalijų. Be to, kuriama sistema aptiktus potencialius gedimus išsaugo duomenų bazėje. Tuo metu saityno programa juos pateikia (kas sekundę atnaujina rodmenis ir potencialių

gedimų sąrašą) įrenginius aptarnaujančiam personalui, nes jie tuo metu įvertina realią situaciją ir priima sprendimą, ką toliau daryti. Konkretnus pavyzdys: yra įrenginys su dvemis nepriklausomais komponentais A ir B (pirmojo darbinė temperatūra yra žymiai didesnė už antrojo). Kažkuriuo laiko momentu B komponentas perkasta ir yra pakeičiamas nauju. Tačiau pakeitus jis vėl kaista. Šiuo atveju siūloma sistema užfiksuos, kad yra reikšminga koreliacija tarp B temperatūros didėjimo ir A temperatūros, ir praneš aptarnaujančiam personalui. Patikrinus įrenginį gali paažiūrėti, kad buvo pažeista izoliacinė medžiaga tarp A ir B komponentų.

B. Iššūkiai

Šitaip analizuojant visus nebūtinai tarpusavyje susijusius duomenis tam tikrose laiko atkarpose nėra garantuojamas priežastinis ryšys. 1965 metais britų medicinos statistikas Austinas Bradfordas Hillas pateikė 9 iki šiol plačiai cituojamus minimalius kriterijus šio ryšio buvimui nustatyti. Sukurtos sistemos analizės algoritmas užtikrina tik keletą iš jų [17]:

- Stiprumas (angl. *strength*) – algoritme naudojamas koreliacijos koeficientas turi būti ne mažesnis už 0,7.
- Pastovumas (angl. *consistency*) – algoritmas lygina su prieš gedimą buvusiais rodmenimis, nes jie yra potencialūs priežasties požymiai.
- Specifiskumas (angl. *specificity*) – daroma prielaida, kad įrenginiai teikia visus įmanomus savo rodmenis.

Bet tai nėra laikoma didėle problema šioje sistemoje, nes tai leidžia aptikti anomalijas, ir galutinį sprendimą priima aptarnaujantis personalas, kuris remdamasis savo dalykinės srities žiniomis gali patvirtinti arba paneigti kelis kitus A. B. Hillo kriterijus (tiketinumą, koherentiškumą, analogiją).

Kitas aptiktas keblumas yra tai, pramonės įmonėms ne visada pakanka vien tik *MTConnect* standarte apibrėžtų duomenų, todėl jos pasinaudoja išplečiamumo galimybėmis ir apsibrėžia savo įrenginiams pritaikytus duomenis. Dėl šios priežasties atsiranda galimybė, kad sukurta sistema nesugebės analizuoti visų įmanomų duomenų, nes ji remiasi tik *MTConnect* specifikaciją.

Trečia iššūkis susijęs su duomenų kiekiu. Athulanas Vijayaraghavanas apžvelgdamas McKinsey ataskaitą apie *didžiuosius duomenis* pateikė savo apytikslius pesimistiškus skaičiavimus, kuriuose matomi labai dideli duomenų kiekiai, kuriuos sugeneruoja *MTConnect* standartu grįsti pramonės įrenginiai:

2 LENTELĖ. MTCONNECT ĮRENGINIŲ SUGENERUOJAMŲ DUOMENŲ KIEKIAI
[18]

	Paprastas stebėjimas		Detalus stebėjimas	
Méginių per sekundę	10	hz	100	hz
Duomenų išrašų mēginyje kiekis	10	išrašų	50	išrašų
Duomenų išrašo dydis	50	bytes	50	bytes
Mégino dydis	500	bytes	2500	bytes
Įrenginio duomenų perdavimo sparta	5	kbps	250	kbps

Dienos duomenų kiekis	432	MB	21600	MB
Duomenų kiekis per metus	158	GB	7889	GB
Metų senumo duomenys saugomi:				
Mažam cechui (10 įrenginių)	2	TB	79	TB
Vidutiniam cechui (30 įrenginių)	5	TB	237	TB
Dideliam cechui (100 įrenginių)	16	TB	789	TB
Pramoninei įmonėi (500 įrenginių)	79	TB	3945	TB
Visa JAV rinka (1,2 mln. įrenginių)	189	PB	9467	TB

Tai buvo žinoma prieš pradedant kurti sistemą, todėl buvo pasirinkta naudoti *Elastic*, kaip dokumentinę duomenų bazę, nes tinka saugoti bet kokios struktūros duomenis ir tikėtina, kad ji veiks itin greitai [19].

V. BANDYMAS

Realizavus sistemą buvo bandoma žiūrėti greitaveiką tikrinant oficialų testavimo įrenginį, kurio agentė pasiekama adresu <http://agent.mtconnect.org>. Sistema buvo įdiegta į nešiojamąjį kompiuterį Acer Aspire V3 (MS Windows 10, Intel Core i7 4702MQ, 16 GB RAM, 1 TB HDD). Per 3-4 valandas iš įrenginio buvo surinkta apie 3 mln. duomenų ir *Elastic* indekso dydis buvo apie 1 GB.

Bandymo metu sistemoje buvo užregistruoti 48 gedimai, buvo analizuojama 15 min. duomenų imtis. Dėl to iš duomenų bazės buvo paimti 188627 naujausi rodmenys ir 5560104 istoriniai rodmenys. Viena šios analizės iteracija užtruko 6 min. 45 sek. (3 lentelė).

3 LENTELĖ. DUOMENŲ ANALIZĖS REZULTATAI

Analizavimo intervalas	15 min.
Praeities gedimai, su kuriais lyginama, skaičius	48
Naujausių rodmenų kiekis	188627
Bendras iš duomenų bazės paimtų istorinių rodmenų kiekis	5560104
Vienos iteracijos vienam įrenginiui trukmė	6 min. 45 sek.
Sunaudota darbinės atminties	Iki 1,8 GB RAM

Tačiau esant tokiai nedidelei imčiai vykdymo trukmė yra per ilga (nepraktiška). Tai galimai atsitiko dėl duomenų formato, perteiklinio duomenų kiekio. Be to, visi trys sistemos moduliai veikė viename fiziniame kompiuteryje ir apkrovė procesorių – todėl jis per daug įkaito ir viso kompiuterio greitaveika sumažėjo.

VI. TOLESNI DARBAI

Tyrimo metu planuojama sumodeliuoti įvairius įrenginio veikimo scenarijus ir atlkti šiuos veiksmus:

- Kadangi kai kurių įrenginių komponentų analizuojami duomenys keičiasi daugiau nei kelis kartus per sekundę, todėl yra planuojama agreguoti duomenis sekundės ir / arba didesnias intervalais.
- Kadangi visi dabartiniai duomenys yra laikomi viename *Elastic* indekse, todėl jis tampa apkrautas ir praranda dalį našumo. Planuojama duomenų analizei naudoti atskirą indeksą, kuriame bus saugomi rodmenys.

- Kadangi sistema saugo visą gautą informaciją ir rodmenų objektas turi įvairių perteiklinių duomenų, todėl planuojama analizei naudoti minimizuotą rodmenų objekto struktūrą.
- Kadangi tarp kaupiamų duomenų yra tekstinių reikšmių, bet sukurta sistema analizuoją tik skaičius, todėl planuojama vykdyti konversiją į skaičius.
- Kadangi analogiškas praeityje buvusiam gedimas gali atsitikti greičiau, todėl analizuojant bus bandoma keisti duomenų imčių mastelį.

Atlikus šiuos veiksmus ir keičiant imtį bus tiriami, kaip keičiasi analizavimo našumas, sunaudojami resursai ir aptiktų potencialių gedimų kiekis.

Tikimasi, kad našumas padidės bent du kartus.

VII. IŠVADOS

Išanalizavus problemą buvo pateiktas siūlomas igyvendintas sprendimas, kuris dėl didelių duomenų kiekių naudoja dokumentinę duomenų bazę ir kuris analizuoją duomenis ieškant koreliacijos.

Siūlomame sprendime buvo identifikuotos ir aprašytos problemas bei pasiūlyti keli sprendimai našumui gerinti keičiant imties intervalus, optimizuojant *Elastic* indeksus, modifikuojant duomenų struktūrą ir mastelį bei vykdant duomenų konvertavimą.

Naudojant siūlomą sprendimą sumažės pramoninių staklių priežiūros kaštai, todėl padidės aptarnaujančio personalo našumas ir jiems bus patogiau prižiūrėti įrenginius.

Kaupiant įrenginių duomenis bus galima peržiūrėti istorinius rodmenis, todėl darbuotojai galės objektyviau įvertinti ir priimti geresnius proceso pakeitimus.

Vykdomi įrenginių duomenų analizė bus galima nuspėti potencialius gedimus, todėl sumažės įrenginio neveikimo laikas ir taisymo kaštai.

Naudojant *Elastic* bus išlaikoma greitaveika didėjant duomenų mastams, todėl sistema veiks efektyviai ilgesnį laiką.

Atlikus tyrimą su planuojamais pakeitimais padidės analizavimo našumas, todėl sistema aptiks greičiau ir daugiau potencialių gedimų.

VIII. LITERATŪRA

- [1] W. Sobel, „MTConnect® Standard. Part 1 - Overview and Protocol,“ MTConnect Institute, 30 9 2014. [Tinkle]. Available: http://static1.squarespace.com/static/54011775e4b0bc1fe0fb8494/t/557f2897e4b04b2acdba80b5/1434396823825/mtc_part_1_overview_w_v1.3.pdf. [Žiūrėta 2016 02 25].
- [2] „MTConnect Standard,“ MTConnect Institute, [Tinkle]. Available: <http://www.mtconnect.org/standard>. [Žiūrėta 2016 02 25].
- [3] Nelson Baxter, Heather De Jesús, „Remote Machinery Monitoring – a Developing Industry,“ 5 2008. [Tinkle]. Available: <http://sandv.com/downloads/0805bax2.pdf>. [Žiūrėta 2014 11 01].
- [4] D. Edstrom, „MTConnect: To Measure Is To Know,“ Virtual Photons Electrons, 2013. [Tinkle]. Available:

- <https://books.google.lt/books?id=Nz4hrkznzR0C>. [Žiūrėta 2016 02 25].
- [5] „Getting Started with MTConnect. Connectivity Guide.“ AMT – The Association For Manufacturing Technology, 2011. [Tinkle]. Available: Getting!Started!with!MTConnect. [Žiūrėta 2016 02 25].
- [6] „MTConnect Challenge.“ U.S. Department of Defense, 2013. [Tinkle]. Available: <https://www.dodmantech.com/Initiatives/MTConnect>. [Žiūrėta 2016 02 25].
- [7] J. Finn, „Integration with Microsoft Visio,“ International TechneGroup Incorporated, 29 4 2013. [Tinkle]. Available: http://s3.amazonaws.com/challengepost/zip_files/production/8102/zip_files/MTConnectResponse-ITI.pdf. [Žiūrėta 2014 11 01].
- [8] V. Pezzullo, „Monitoring and Control of Chatter Conditions with MTConnect,“ Clemson University, 31 5 2013. [Tinkle]. Available: http://s3.amazonaws.com/challengepost/zip_files/production/8428/zip_files/MTConnect_Challenge_1_Valerie_Pezzullo.pdf. [Žiūrėta 2014 11 01].
- [9] R. Sivalingam, „Google Glass and the 4th Industrial Revolution,“ 1 9 2014. [Tinkle]. Available: <http://www.academia.edu/8306128/>. [Žiūrėta 2014 11 01].
- [10] Byeong Eon Lee; John L. Michaloski; Frederick M. Proctor; Sid Venkatesh; Nils Bengtsson, „MTConnect-Based Kaizen for Machine Tool Processes,“ Computers and Information in Engineering Conference, 15 8 2010. [Tinkle]. Available: http://www.nist.gov/customcf/get_pdf.cfm?pub_id=905591. [Žiūrėta 2014 11 01].
- [11] Vijayaraghavan, Athulan; Sobel, Will; Fox, Armando; Dornfeld, David; & Warndorf, Paul, „Improving Machine Tool Interoperability Using Standardized Interface Protocols: MT Connect,“ UC Berkeley: Laboratory for Manufacturing and Sustainability, 23 6 2008. [Tinkle]. Available: <http://escholarship.org/uc/item/4zs976kx>. [Žiūrėta 2014 11 01].
- [12] Sri Atluru, Amit Deshpande, „Statistical Process Monitoring With MTConnect,“ Proceedings of the ASME 2012 International Manufacturing Science and Engineering Conference, 4 7 2012. [Tinkle]. Available: http://www.researchgate.net/publication/254201445_STATISTICA_L_PROCESS_MONITORING_WITH_MTCONNECT. [Žiūrėta 2014 11 01].
- [13] John L. Michaloski; Byeong Eon Lee; Frederick M. Proctor; Sid Venkatesh, „Web-enabled Real-time Quality Feedback for Factory Systems using MTConnect,“ ASME 2012 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, 12 8 2012. [Tinkle]. Available: http://www.nist.gov/customcf/get_pdf.cfm?pub_id=911323. [Žiūrėta 2014 11 01].
- [14] Sri Atluru, Amit Deshpande, Sam Huang, Ron Pieper, „PneuViz: MTConnect Compliant Compressed Air Monitoring Application,“ SME 2012 International Manufacturing Science and Engineering Conference, 4 6 2012. [Tinkle]. Available: <http://www.home.amit-deshpande.com/mysite/Papers/MSEC2012Pneu.pdf>. [Žiūrėta 2014 11 01].
- [15] John L. Michaloski; Byeong Eon Lee; Frederick M. Proctor; Sid Venkatesh; Sidney Ly, „Quantifying the Performance of MT-Connect in a Distributed Manufacturing Environment,“ ASME 2009 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, 30 8 2009. [Tinkle]. Available: http://www.nist.gov/customcf/get_pdf.cfm?pub_id=903259. [Žiūrėta 2014 11 01].
- [16] Ben Edrington, Bingyan Zhao, Adam Hansel, Masahiko Mori, Makoto Fujishima, „Machine Monitoring System Based on MTConnect Technology,“ 3rd International Conference on Through-life Engineering Services, 1 4 2014. [Tinkle]. Available: <http://www.sciencedirect.com/science/article/pii/S2212827114009664>. [Žiūrėta 2014 11 01].
- [17] A. B. Hill, „The Environment and Disease: Association or Causation?,“ *Proceedings of the Royal Society of Medicine*, nr. 58, pp. 295-300, 1965.
- [18] A. Vijayaraghavan, „McKinsey on (Manufacturing) Big Data: Part 1 - How Much Data?,“ 13 02 2012. [Tinkle]. Available: <http://www.systeminsights.com/blog/2012/2/13/mckinsey-on-manufacturing-big-data-part-1-how-much-data>. [Žiūrėta 2016 02 25].
- [19] Chris Golledge, Anup Nair, „Introducing NoSQL Capabilities. A Technical White Paper,“ IBM Corporation, 2013. [Tinkle]. Available: https://www.ibm.com/developerworks/community/blogs/idssteam/resource/IFMX-12.10.xC2-WP-Overall_20131018.pdf?lang=en. [Žiūrėta 2016 02 25].

Mobile identification to VPN system using crypto processor and NFC

A. Nauckūnas, E. Sakalauskas

Kaunas University of Technology

Faculty of Mathematics and natural sciences

Kaunas, Lithuania

arvydas.nauckunas@gmail.com

Abstract — Mobile VPN identification system is presented that uses microSD based Crypto processor and NFC. Whole system includes three main communicating devices: remote VPN server, local computer and mobile phone with microSD based crypto processor. The overall identification system's security is estimated.

Keywords — VPN, NFC, OpenVPN, Cryptoki, PKI, Crypto processor.

I. INTRODUCTION

Mobile phone has become part of every person's life. Now mobile phones can replace many devices while in the same time providing more features to the users. Phone now acts as a photo camera, calculator, video camera and other devices.

In today's age of information there are a lot of different computerized systems where users have to remember their user names and passwords in order to work with and maintain those systems. With increasing amount of information and its sensitivity and a need to comply with security requirements, mobile identification systems are becoming more relevant to every person. Passwords become even longer to comply with growing security standards and they need to be frequently changed.

Public-private key authentication method is being used increasingly. This provides improved security compared to password authentication due to the fact that a key length is usually much longer than password and does not represent meaningful words or patterns therefore it is impossible to guess or brute-force with currently known hardware. Unfortunately using PKI brings its own challenges such as safe key storage and portability combined with ease of use.

Computers that are communicating with the private networks are usually connected to these networks via an unsecure channel. It could be either the Internet or a special Internet Service provider link. In order to reach these networks in a secure manner VPN connections are established. This ensures data encryption and integrity. To establish this secure channel password authentication or more secure private-public key authentication is used. This brings forth challenges how to use mobile phone as a private key storage and how to establish VPN for local computers.

II. PROPOSED MOBILE VPN IDENTIFICATION SOLUTION

One of the most popular ways to handle a private key is to store it on a local computer. But this approach restricts usage with only one computer. Portability problem is usually addressed by carrying a private key in USB thumb drive. As an additional security measure password can be set on a private key file to prevent its usage without knowing the password. This protection usually can be cracked by using dictionary or brute force attacks.

Introducing private key to the computer systems where it can be directly read adds risk of it being stolen.

The proposed solution is a way to solve private key portability and usage for VPN identification security issues.

A. Solution components

Components for the solution were selected based on how widely spread the technology is and security benefit it brings.

- NFC enabled smartphone with a microSD card slot and Android 4.0.3 or higher. It was selected because of a wide spread and Android version 4.0.1 and lower has already faced NFC related security concerns. Also currently Android takes most of the market share of all smartphone operating systems. Where version 4.0.3 and higher make up more than 90% of Android OS;
- microSD based Crypto processor which complies with PKCS#11 standard. Hardware solution for private key storage was selected since it protects private key without an ability to extract it and only provides one entry point which is PIN protected for crypto function operations. Selected microSD cards also provides storage as standard cards without crypto processor;
- NFC reader built-in or attached to a local computer. NFC short range communication minimizes risk of various attacks such man in the middle. Does not require special setup before usage;

- Android application that handles NFC communication and interaction with microSD card functions. Through the application data is passed to crypto functions on microSD card. It also acts as VPN profile storage that contain VPN parameters and handles NFC communication;
- Agent application on a local Windows computer that handles communication between OpenVPN client and NFC;
- Remote OpenVPN server [1] as part of complete solution provides access to secured network resources over VPN tunnel based on user identification.

This solution can also be used with other smartphones as well if there is vendor provided library for crypto microSD card for particular OS.

Mobile devices with iOS do not have a slot for external microSD card. Therefore they are unable to make best use of this solution. Tested crypto microSD card vendor enables to use this proposed solution by connecting microSD card reader via iOS device lightning port.

B. VPN identification

Mobile identification in this solution taps into existing TLS authentication protocol and OpenVPN “management” and “management-external-key” functions [2].

Main VPN identification steps in Fig. 1 of the proposed solution:

1. Non-sensitive data transfer from mobile phone to a local computer;
2. Initial TLS handshake until TLS authentication;
3. SHA1 (160 bits) and MD5 (128 bits) hash calculation according to TLS authentication protocol. Transferring a combined hash with length of 288 bits to Android mobile phone application for hash signing with a private key within crypto processor based microSD card;
4. Finalizing TLS handshake.

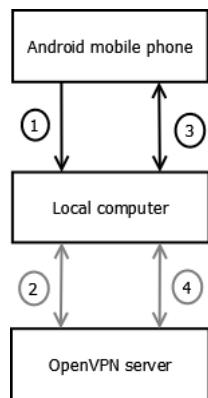


Fig. 1. VPN identification steps of the proposed solution

Identification steps 2 and 4 in Fig. 1 are handled by OpenVPN therefore they will not be discussed further.

User initiates identification by selecting profile in Android application which relates OpenVPN configuration parameters, CA and client certificates and identifier for a private key that is stored on PKCS#11 microSD. After that user PIN code has to be entered for interaction with objects in microSD. If code is correct then user touches mobile phone to NFC reader. From this point Android application and agent service installed on a local computer handles identification process.

Android application sends VPN parameters, CA and client certificates to a local computer via NFC protocol. Agent service on a local computer employs received information to initiate OpenVPN TLS connection. This covers identification step 1.

During Step 3 according to TLS protocol SHA1 [3] (160 bits) and MD5 [4] (128 bits) hash values of all previous communication are calculated. So the total value of 288 bits has to be signed within 60 seconds with a private key corresponding to previously provided client certificate. Agent program sends this value to Android application which passes it to microSD along with a private key identifier for signing operation. Signature is then sent back to agent program via NFC and passed on to OpenVPN server for authentication.

If hash value was signed with correct private key and signature was submitted to OpenVPN server within 60 seconds of TLS initialization then authentication is successful and VPN connection gets established.

Once VPN connection is established mobile phone can be removed from NFC reader since it is used only to provide configuration and user authentication at initial stages. There is no further interaction with mobile or crypto microSD card after hash signature is sent to a local computer.

In case network connection is lost and VPN tunnel disconnects then mobile identification steps have to be repeated in order to establish new VPN connection.

III. SYSTEM SECURITY

Whole system security is based on how each component is secured and where sensitive data is located and used.

System is built from three main communicating devices:

- Android mobile phone – used as OpenVPN client CA and client certificate holder. Private key is stored as an object on microSD card that supports PKCS#11;
- Local computer – intended OpenVPN client that will establish VPN connection and gain access to remote resources;
- Remote OpenVPN server – used to provide access to remote networks or resources. It can be located on internal network without any access from Internet to secure private network resources or connection can be established over Internet.

A. PKCS#11 microSD card

PKCS#11 microSD card cryptographic functions in the system are only reachable through provided Android library which uses microSD read and write commands to invoke special crypto processor functions.

Objects contained in the card can only be manipulated with when API authenticates against the card with initially set PIN code. Without it any PKCS#11 operations with the card objects (private keys, certificates) are prohibited.

Main function used for this solution is RSA PKCS #1 v1.5 signature RSASP1 (K, m), where input parameters are as follows [5]:

K - RSA private key, where K has one of the following forms:

- pair (n, d)
- quintuple $(p, q, dP, dQ, qInv)$

m - message representative, an integer between 0 and $n-1$

Output is a signature s represented by an integer between 0 and $n-1$. Assumption of the calculation is that private key K is valid.

1. If the message representative m is not between 0 and $n-1$ then message is too long and cannot be signed.

2. If the first form (n, d) of K is used:

2.1 Let $s = m^d \bmod n$.

Else, if the second form $(p, q, dP, dQ, qInv)$ of K is used:

2.2 Let $s_1 = m^{dP} \bmod p$.

2.3 Let $s_2 = m^{dQ} \bmod q$.

2.4 Let $h = qInv (s_1 - s_2) \bmod p$.

2.5 Let $s = s_2 + hq$.

3. Output s .

RSA signature length depends on a private key size. For private key which size is 1024 bits the signature size is 128 bytes and for 2048 bit – 256 bytes.

Timings were measured when signing with different key sizes in crypto processor. Signing, including user authentication against microSD card and other related operations, took around 2 seconds in total. In this particular crypto processor realization there was no difference in timings when signing with different key sizes.

B. Android mobile phone

System security on mobile phone relies on operating systems security. Since private key is not retrievable from microSD card and without knowledge of the PIN code it is not possible to use any PKCS#11 functions with private key object.

The only way PIN of microSD card can be obtained is if user gives this information to someone or mobile phone keyboard button presses are being monitored. More difficult

approach to acquiring PIN code would be to monitor all system calls sent to microSD card.

Prevention from technological aspect of stealing PIN code can be mitigated by carefully selecting applications that mobile phone user is installing and what permissions they require. As well as keep mobile phone running without phone user acquiring elevated user privileges. Getting root permissions on mobile phone widens attack surface and introduces more possible security threats with bigger impact due to full access to operating system and its processes.

C. NFC channel

NFC communication is possible only between two devices. In this case it is between mobile phone and Windows computer with NFC reader. Reader can be built-in or connected as peripheral. Although theoretical working distance of NFC is less than 10cm after number of tests with mobile phones and several NFC readers the actual measured distance is no more than 1.5cm. And that can only be achieved if mobile phone is held at certain angle against NFC reader. This distance limits possible information theft or man-in-the-middle attacks unless reader itself has been physically tampered with.

Even if NFC communication would be eavesdropped in this proposed solution NFC channel handles only data that are not sensitive and cannot be used for anything except for currently ongoing authentication.

Data transferred via NFC is:

- CA and Client certificates – can be publicly viewed;
- OpenVPN parameters – file containing remote OpenVPN server IP address and port number. This information can be obtained by doing port scan;
- Hash values – related only to current TLS connection and cannot be used elsewhere. Even with current connection they become invalid after 60 seconds;
- Signature – related only to current TLS connection and cannot be used elsewhere. Even with current connection they become invalid after 60 seconds.

Data transfer rates were tested with different phones and NFC readers. When phones were in close enough proximity to establish NFC communication the data rates were always according to standard specifying data rate of 424 kbit/s. Actual transfer speed measured was around 40KB/s. Amount of data that needs to be transferred during VPN identification mainly depends on certificate and signature sizes. OpenVPN configuration parameters used are up to size of 1KB. With tested key length of 2048 bits whole NFC data transfer took less than 0.2 seconds.

D. Local computer

A local computer has to be secure enough for particular VPN connections that are to be established. No sensitive data related to VPN connection is stored on a local computer. During VPN tunnel setup phase local computer handles only

information that is transferred via NFC. In general if more than one user is using local computer some files can be inspected by another user. After successful establishment of VPN tunnel all configuration data are removed by running agent program. Sensitive data such as a private key is never present on local computer. If unauthorized user gets access to computer he will not be able to reach remote network resources over VPN.

Agent program only allows initiate VPN connection when user is logged in with his user profile. It is not possible to establish connection if user is in lock screen. To avoid someone else establishing VPN connection without knowledge of current user working with the computer agent service shows notification when VPN connection is being established. Only single OpenVPN connection instance is allowed at one time.

IV. OVERALL SECURITY ANALYSIS

Main focus of proposed system security is a way how a private key is accessed and ability to use it with cryptographic functions. Retrieval of a private key object from microSD card is not possible therefore it is important to prevent its usage with cryptographic functions.

Standard way to use cryptographic processor functions is via library built into Android application. Through that library user presented PIN code is passed on to microSD card crypto processor for authentication.

Maintaining overall operating systems security is best way to keep access to crypto processor secured. Without privileged access to operating systems privileged “root” user no other application or person will be able to listen system calls to microSD card. This also prevents from unauthorized inspection of keyboard input.

NFC has several defined standards that cover NFC channel security. ECMA [6] has released standard ECMA-385 [7] that defines available NFC security services and protocol NFC-SEC. While extending protocol with ECMA-386 [8] adding ECDH and AES support. These standards use key agreement mechanism based on ISO/IEC 11770-3 [9].

Another standardization body NFC Forum has released version 1.3 of the LLCP specification [10] that adds an unauthenticated secure data transport option to ensure privacy and confidentiality of messages exchanged between peer devices.

While these standards exist to ensure NFC channel security between two peer devices they are not yet implemented on mobile phones. This relates to the fact that due to very short NFC communication range it is unlikely that someone will be able to abuse NFC data transfers therefore NFC security standards are not yet widely implemented. In proposed solution no sensitive data are being transferred.

V. CONCLUSION

The proposed VPN identification solution brings security and portability benefits with minimal cost. Adoption of crypto processor based microSD card secures private keys and all related cryptographic operations.

Even if mobile phone or cryptographic card would be lost or stolen a private key object and related functions would not be accessible without knowledge of the PIN code. One could try to brute force PIN code but it is long enough to give enough time for the owner to invalidate all access and revoke certificates related to the stolen keys. Cryptographic cards do not differ from standard cards in visual appearance or from operating system perspective. Cost of same size microSD cards with crypto processor are only marginally higher.

Most of the interaction during VPN identification process with the proposed solution has been automated therefore requiring less user interaction while providing improved security and portability benefits.

The proposed solution is used to authenticate local computer against a remote OpenVPN server and create VPN connection with the help of smartphone. Similar solutions that use crypto microSD card authenticate not a local computer but smartphone itself in order to create VPN tunnel between phone and remote VPN server.

VI. ABBREVIATIONS AND ACRONYMS

NFC – near field communication;

API – application programming interface;

PIN – private identification number;

PKCS#11 - standard defines a platform-independent API to cryptographic tokens;

TLS – transport layer security;

SHA1 – secure hashing algorithm version 1;

USB – universal serial bus;

VPN – virtual private network;

IP address – numerical label assigned to each device participating in a computer network that uses the Internet Protocol for communication;

port – endpoint of communication in an operating system;

AES – Advanced Encryption Standard, symmetric encryption protocol;

ECDH – Elliptic curve Diffie–Hellman key agreement protocol;

CA – certificate authority.

REFERENCES

- [1] OpenVPN [Online]. Available: <https://openvpn.net/>
- [2] OpenVPN. Management interface [Online]. Available: <https://openvpn.net/index.php/open-source/documentation/miscellaneous/79-management-interface.html>
- [3] US Secure Hash Algorithm 1 (SHA1) [Online]. Available: <https://tools.ietf.org/html/rfc3174>
- [4] The MD5 Message-Digest Algorithm [Online]. Available: <http://tools.ietf.org/rfc/rfc1321>
- [5] PKCS #1: RSA Cryptography Specifications Version 2.0. RSASPI [Online]. Available: <https://tools.ietf.org/html/rfc2437#section-5.2.1>

- [6] ECMA International [Online]. Available: <http://www.ecma-international.org/>
- [7] NFC-SEC; NFCIP-1 Security Services and Protocol [Online]. Available: <http://www.ecma-international.org/publications/files/ECMA-ST/ECMA-385.pdf>
- [8] NFC-SEC-01: NFC-SEC Cryptography Standard using ECDH and AES [Onlinel. Available: <http://www.ecma-international.org/publications/files/ECMA-ST/ECMA-386.pdf>
- [9] ISO/IEC 11770-3:2015. Information technology -- Security techniques - - Key management -- Part 3: Mechanisms using asymmetric techniques [Online]. Available: http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=60237
- [10] NFC Forum Technical Specifications [Online]. Available: <http://nfc-forum.org/our-work/specifications-and-application-documents/specifications/nfc-forum-technical-specifications/>

Informacinės technologijos : konferencijos pranešimų medžiaga / Kauno technologijos universitetas, Vytauto Didžiojo universitetas, Vilniaus universitetas ; atsakingoji redaktorė Erinija Pranckevičienė. – [1999]- . – Kaunas : Technologija, 1999- [2016] : XXI tarpuniversitetinės tarptautinės magistrantų ir doktorantų konferencijos „Informacinė visuomenė ir universitetinės studijos“ (IVUS 2016) medžiaga : 2016 m. balandžio 28 d., Kaunas, Lietuva = [2016] : proceedings of the XXI international Master and PhD conference “Information society and university studies” (IVUS 2016) : 28th April, 2016, Kaunas, Lithuania / Vytauto Didžiojo universitetas, Kauno technologijos universitetas, Vilniaus universiteto Kauno humanitarinis fakultetas ; redakcinė kolegija: Tomas Krilavičius ... [et al.]. – Kaunas : Vytauto Didžiojo universitetas, 2016. – 77 p. : iliustr. – Bibliografija straipsnių gale.

ISSN 2029-249X (spausdintas)
ISSN 2029-4824 (internetinis)

INFORMACINĖS TECHNOLOGIJOS

XXI tarpuniversitetinės tarptautinės magistrantų ir doktorantų konferencijos „Informacinė visuomenė ir universitetinės studijos“ (IVUS 2016) medžiaga, 2016 m. balandžio 28 d., Kaunas, Lietuva

INFORMATION TECHNOLOGY

Proceedings of the XXI International Master and PhD Students Conference “Information Society and University Studies” (IVUS 2016), 28th April 2016, Kaunas, Lithuania

Editor Audrius Varoneckas

Approved for printing 2016 04 25. A run of 24 copies. (No of P. 77)
Order No. K16-023.

Published: Vytautas Magnus University, K. Donelaičio g. 58, LT-44248 Kaunas
www.vdu.lt | leidyba@bibl.vdu.lt

Redaktorius Audrius Varoneckas

Pasirašyta spausdinti 2016 04 25. Tiražas 24 egz. 4,85 leidyb. apsk. l. (77 p.)
Užsakymo Nr. K16-023.

Įšleido: Vytauto Didžiojo universitetas, K. Donelaičio g. 58, LT-44248 Kaunas
www.vdu.lt | leidyba@bibl.vdu.lt