ISSN 2710-0766
DOI 10.31891/CSIT

THE INTERNATIONAL SCIENTIFIC JOURNAL

COMPUTER SYSTEMS AND INFORMATION TECHNOLOGIES

No 4-2022

МІЖНАРОДНИЙ НАУКОВИЙ ЖУРНАЛ

КОМП'ЮТЕРНІ СИСТЕМИ ТА ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ

2022

COMPUTER SYSTEMS AND INFORMATION TECHNOLOGIES

INTERNATIONAL SCIENTIFIC JOURNAL

Published since 2020 year *Four time a year*

Khmelnitskyi, 2022, № 4 (9)

Establishers: **Khmelnytskyi National University** (Ukraine)
Associated establisher: **Institute of Information Technologies** (Slovakia)
National Library og Ukraine named after V.I.Vernadsky http://nbuv.gov.ua/j-tit/csit

The journal is included in scientometric databases:
Index Copernicus https://journals.indexcopernicus.com/search/details?id=69998&lang=en
Google Scholar https://scholar.google.com.ua/citations?hl=uk&user=HW1XpMsAAAAJ
CrossRef http://doi.org/10.31891/CSIT

Editors **Hovorushchenko T.,** Doctor of engineering sciences, Professor, Head of the Department of Computer Engineering and Information Systems Khmelnytskyi National University (Ukraine)

Head editorial board **Savenko, O.,** Doctor of engineering sciences, Professor of the Department of Computer Engineering and Information Systems, Dean of the Faculty of Information Technologies, Khmelnytskyi National University (Ukraine)

Executive secretary **Lysenko S.,** Doctor of engineering sciences, Professor of the Department of Computer Engineering and Information Systems Department, Khmelnytskyi National University (Ukraine)

Editorial board:

Hovorushchenko T., Doctor of engineering sciences (Khmelnytskyi, Ukraine), **Savenko O.,** Doctor of engineering sciences (Khmelnytskyi, Ukraine), **Barmak O.,** Doctor of engineering science (Khmelnytskyi, Ukraine), **Lysenko S.,** Doctor of engineering sciences (Khmelnytskyi, Ukraine), **Peter Popov,** PhD (London, Great Britain), **Piotr Gaj,** Dr hab inż (Gliwice, Poland), **Anatolii Gorbenko**, DrSc, Professor (Leeds, Great Britain), **Andrzej Kotyra,** Dr hab inż, Professor (Lublin, Poland), **Andrzej Kwiecień,** Dr hab inż, Professor (Gliwice, Poland), **George Markowsky,** PhD in Mathematics, Professor of Computer Science (Missouri, USA), **Sergii Babichev**, DrSc, Professor (Ústí nad Labem, Czech Republic), **Krak Iu**., Doctor of mathematics and physics sciences (Kyiv, Ukraine), **Yatskiv V.,** Doctor of engineering sciences (Ternopil, Ukraine), **Pastukh O.,** Doctor of engineering sciences (Ternopil, Ukraine), **Romankevich V.**, Doctor of engineering sciences (Kyiv, Ukraine), **Sachenko A.,** Doctor of engineering sciences (Ternopil, Ukraine), **Korobchynskyi M.,** Doctor of engineering sciences (Kyiv, Ukraine), **Bisikalo O.,** Doctor of engineering sciences (Vinnytsia, Ukraine), **Maevsky D.,** Doctor of engineering sciences (Odesa, Ukraine), **Zharikova M.,** Doctor of engineering sciences (Kherson, Ukraine), **Sherstjuk V.,** Doctor of engineering sciences (Kherson, Ukraine), **Berezsky O.,** Doctor of engineering sciences (Ternopil, Ukraine), **Yakovyna V.**, Doctor of engineering sciences (Lviv, Ukraine), **Lupenko S.,** Doctor of engineering sciences (Ternopil, Ukraine), **Shilo G.,** Doctor of engineering sciences (Zaporijjya, Ukraine), **Bobrovnikova K.,** PhD (Khmelnytskyi, Ukraine), **Nicheporuk A.,** PhD (Khmelnytskyi, Ukraine), **Hnatchuk Y.,** PhD (Khmelnytskyi, Ukraine), **Medzatyi D.**, PhD (Khmelnytskyi, Ukraine), **Perepelitsyn A.,** PhD (Kharkiv, Ukraine), **Illiashenko O.,** PhD (Kharkiv, Ukraine), **Izonin I.,** PhD (Lviv, Ukraine), **Horiashchenko S.,** PhD (Khmelnytskyi, Ukraine), **Boyarchuk A.,** PhD (Kharkiv, Ukraine), **Pavlova O.,** PhD (Khmelnytskyi, Ukraine)

Technical editor **Kravchyk Yu,** PhD.

Recommended for publication by the decision of the Academic Council of Khmelnytskyi National University, protocol № 6 from 29.12.2022

Editorial board address: International scientific journal "Computer Systems and Information Technologies", Khmelnytskyi National University, Institutska str. 11, Khmelnytskyi, 29016, Ukraine

☎ **(0382) 67-51-08**

e-mail: csit.khnu@gmail.com

web: http://csitjournal.khmnu.edu.ua/index.php/csit
http://lib.khnu.km.ua/csit_khnu.htm

Registered by the Ministry of Justice of Ukraine
Certificate of state registration of the print media
Series KB № 24924-14864PR dated 12.07.2021

© Khmelnytskyi National University, 2022
© Editorial board "Computer Systems and Information Technologies", 2022

КОМП'ЮТЕРНІ СИСТЕМИ ТА ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ

МІЖНАРОДНИЙ НАУКОВИЙ ЖУРНАЛ

Засновано в 2020 р. ***Виходить 4 рази на рік***

Хмельницький, 2022, № 4 (9)

Засновник і видавець: **Хмельницький національний університет** (Україна)
Асоційований співзасновник: **Інститут інформаційних технологій** (Словаччина)

Наукова бібліотека України ім. В.І. Вернадського http://nbuv.gov.ua/j-tit/csit

Журнал включено до наукометричних баз:
Index Copernicus https://journals.indexcopernicus.com/search/details?id=69998&lang=en
Google Scholar https://scholar.google.com.ua/citations?hl=uk&user=HW1XpMsAAAAJ
CrossRef http://doi.org/10.31891/CSIT

Головний редактор **Говорущенко Т. О.,** д. т. н., професор, завідувач кафедри комп'ютерної інженерії та інформаційних систем Хмельницького національного університету

Заступник головного редактора. Голова редакційної колегії **Савенко О. С.,** д. т. н., професор, професор кафедри комп'ютерної інженерії та інформаційних систем, декан факультету інформаційних технологій Хмельницького національного університету

Відповідальний секретар **Лисенко С. М.,** д. т. н., професор, професор кафедри комп'ютерної інженерії та інформаційних систем Хмельницького національного університету

Ч л е н и р е д к о л е г і ї

Говорущенко Т.О., д. т. н., **Савенко О.С.,** д. т. н., **Бармак О. В.** д. т. н., **Лисенко С.М.** д. т. н., **Пітер Попов,** доктор філософії (Лондон, Велика Британія), **Пьотр Гай,** д. т. н. (Глівіце, Польща), **Анатолій Горбенко**, д. т. н. (Лідс, Велика Британія), **Анджей Котира,** д .т. н. (Люблін, Польща), **Анджей Квецен,** д. т. н. (Глівіце, Польща), **Джордж Марковський,** к. ф-м. н. (Міссурі, США), **Сергій Бабічев** (Усті над Лабем, Чехія), **Крак Ю.В.** д. ф.-м. н., **Яцків В. В.** д. т. н., **Пастух О.А.,** д. т. н., **Романкевич В.О.,** д. т. н., **Саченко А.О.,** д. т. н., **Коробчинський М.В.,** д. т. н., **Бісікало О.В.,** д. т. н., **Маєвський Д.А.,** д. т. н., **Жарікова М.В.,** д. т. н., **Шерстюк В.Г.,** д. т. н., **Березький О.М.,** д. т. н., **Яковина В.С.,** д. т. н., **Лупенко С.А.,** д. т. н., **Шило Г.М.,** д. т. н., **Бобровнікова К.Ю.,** к. т. н., **Нічепорук А.О.,** к. т. н., **Гнатчук Є.Г.,** к. т. н., **Медзатий Д.М.,** к. т. н., **Перепелицин А.Є.,** к. т. н., **Ілляшенко О.О.,** к. т. н., **Ізонін І.В.,** к. т. н., **Горященко С.Л.,** к. т. н., **Боярчук А.В.,** к. т. н., **Павлова О.О.,** д.ф.

Технічний редактор **Кравчик Ю. В.,** к. е. н., доцент

Рекомендовано до друку рішенням Вченої ради Хмельницького національного університету, протокол № 6 від 29.12.2022

Адреса редакції: **Україна, 29016,**
м. Хмельницький, вул. Інститутська, 11,
Хмельницький національний університет
редакція журналу "Комп'ютерні системи та інформаційні технології"
☎ **(0382) 67-51-08**
e-mail: csit.khnu@gmail.com
web: http://csitjournal.khmnu.edu.ua/index.php/csit
http://lib.khnu.km.ua/csit_khnu.htm

Зареєстровано Міністерством юстиції України
Свідоцтво про державну реєстрацію друкованого засобу масової інформації
Серія КВ № 24924-14864ПР від 12 липня 2021 року

© Хмельницький національний університет, 2022
© Редакція журналу "Комп'ютерні системи та інформаційні технології", 2022

CONTENTS

ЗМІСТ

UDC 004.91
https://doi.org/10.31891/csit-2022-4-1

IRYNA ZASORNOVA, SERGII LYSENKO, OLEKSANDR ZASORNOV
Khmelnytskyi National University

CHOOSING SCRUM OR KANBAN METHODOLOGY FOR PROJECT MANAGEMENT IN IT COMPANIES

Taking into account the conditions of the functioning of modern business, it is necessary to take into account the constant changes that occur in the process of developing projects in IT companies. Therefore, when creating projects, there is a need for a well-founded choice of project management methodologies, which will allow to quickly solve problems that arise in the process of their creation and will provide an opportunity to ensure high efficiency of project-oriented activities of companies.

The conducted analysis of the methodologies showed that today there are a sufficient number of them, among which there are rigid and flexible ones. Among the hard methodologies, as Waterfall model, V-model, Incremental model, Spiral model, and Iterative model are known. Among flexible methodologies, the Agile model, as well as Scrum and Kanban are most often used. The following flexible methodologies are also known: Lean, eXtreme Programming (XP), Rational Unified Process (RUP), Dynamic Systems Development Model (DSDM), Rapid Application Development (RAD), and Extreme Programming (XP).

It has been determined that most researchers prove in their studies that agile methodologies are more widespread nowadays and are better adapted to the conditions of rapid changes in projects. The only exceptions are those methodologies that apply to certain industries, such as medical, aviation, and others. Such fields involve the use of only brutal methodologies.

The article examines the modern Agile approach to project management, its difference from the traditional approach of the rigid Waterfall methodology or others; the scope of application of Agile; a short story; the advantages and disadvantages of implementation and its use.

The article provides a comparative analysis of the most popular today's flexible Scrum and Kanban methodologies. The roles that are indicated in projects in Scrum and Kanban are considered. The events performed by the teams are also analyzed in more detail.

The article elaborates a block diagram for a justified choice of Scrum or Kanban methodology. The article is based on research on the international and Ukrainian markets of IT companies.

The direction of future research of the authors is to conduct a more detailed analysis and comparison of rigid and flexible methodologies with the aim of using combined forms of project management.

Keywords: IT companies, rigid and flexible methodologies, Waterfall model, V-model, Incremental model, Spiral model, Iterative model, Agile model, Scrum and Kanban, roles in the project, Product Owner, Scrum team, Scrum master, Sprint Planning, Daily Scrum, Sprint Review, Sprint Retrospective, backlog, block diagram, comparative analysis, justified choice of methodology.

ІРИНА ЗАСОРНОВА, СЕРГІЙ ЛИСЕНКО, ОЛЕКСАНДР ЗАСОРНОВ
Хмельницький національний університет

ВИБІР МЕТОДОЛОГІЇ SCRUM АБО KANBAN ДЛЯ УПРАВЛІННЯ ПРОЄКТАМИ В ІТ КОМПАНІЯХ

Враховуючи умови функціонування сучасного бізнесу потрібно враховувати постійні зміни, які виникають в процесі розробки проєктів в ІТ компаніях. Тому, при створенні проєктів виникає необхідність обґрунтованого вибору методологій управління проєктами, що дозволить швидко вирішувати проблеми, які виникають у процесі їх створення, та дасть можливість забезпечити високу ефективність проєктно-орієнтованої діяльності компаній.

Проведений аналіз методологій показав, що на сьогодні їх існує достатня кількість, серед яких є жорсткі та гнучкі. Було визначено, що більшість дослідників доводять у своїх працях, що гнучкі методології є більш широко розповсюдженими в наш час і краще адаптуються до умов швидких змін у проєктах. Виключення становлять лише ті методології, які застосовуються для певних галузей, таких як медична, авіаційна та інші. Такі галузі передбачають використання тільки жорстиких методологій.

У статті розглядається відносно сучасний підхід Agile до управління проєктами, його відмінність від традиційного підходу жорсткої методології, наприклад Waterfall, або інших; область застосування Agile; коротка історія; переваги та недоліки впровадження та його використання.

У статті виконано порівняльний аналіз найпопулярніших на сьогодні гнучких методологій Scrum і Kanban. Розроблено блок-схему для обґрунтованого вибору методології Scrum або Kanban. Стаття базується на дослідженнях міжнародного та українського ринку ІТ компаній.

Напрямком майбутніх досліджень авторів є проведення більш детального аналізу та порівняння жорстких і гнучких методологій з метою використання комбінованих форм управління проєктами.

Ключові слова: ІТ компанії, жорсткі та гнучкі методології, Waterfall модель, V- модель, Incremental модель, Spiral модель, Iterative модель, Agile модель, Scrum і Kanban, ролі у проєкті, Product Owner, Scrum команда, Scrum майстер, Sprint Planning, Daily Scrum, Sprint Review, Sprint Retrospective, backlog, блок-схема, порівняльний аналіз, обґрунтований вибір методології.

Introduction

The operating conditions of modern business in the world market are characterized by constant changes. In recent years, these changes have become even more dynamic. Reasonable choice of project management methodologies will allow solving problems that arise in the process of their creation and will provide an opportunity to ensure high efficiency of project-oriented activities of companies.

The variety of concepts and a large number of different applications for the implementation of team projects in IT companies, in particular for the development of software, complicates the choice of the necessary methodology. Therefore, a scientifically based choice of methodology for obtaining high-quality development products in a timely manner is an urgent task. At the same time, the following information must be taken into account: project type, customer requirements, resources, deadlines, tools, team composition, and others.

We will analyze the methodologies used by IT companies. So, today there are different methodologies for managing the implementation of projects in companies. The choice of methodology depends on the type of company and the project it is developing. In modern companies, project management methodologies are usually divided into rigid and flexible, sometimes they use their mixed form [1, 2]. Rigid methodologies include:

– Waterfall model – in this methodology, the stages depend on each other and the next one begins when the previous one is completed, thus forming a progressive (cascading) forward movement. Teams of different stages do not communicate with each other, each team is clearly responsible for its own stage. Advantages: all stages of the project are performed in a strict sequence; the strictness of the stages allows you to plan the deadlines for the completion of all works and the corresponding resources (money and human); requirements remain the same throughout the cycle. Disadvantages: difficulties in formulating clear requirements and the impossibility of changing them; testing begins only in the middle of project development; until the development process is complete, users cannot be sure whether the product being developed is of good quality;

– V-model – involves dividing the project into parts (stages, iterations) and passing the stages of the life cycle on each of them. The use of an iterative model reduces risks and makes it possible to complete the development at the end of each iteration, the set of stages forms the final result. Advantages: strict phasing; minimization of risks and elimination of potential problems due to the fact that testing appears in the early stages; improved time management. Disadvantages: inability to adapt to the customer's changed requirements; long development time (sometimes lasting up to several years) leads to the fact that the product may not be needed by the customer, as his needs change; there are no actions embedded in the risk analysis;

– Incremental model – the cycle is divided into smaller modules that are easy to create. Each module goes through requirements definition, design, coding, implementation and testing phases. The development procedure according to the incremental model involves the release of the product in the first major stage in the basic functionality, and then the sequential addition of new functions. The process continues until a complete system is created. Advantages: the customer can give feedback on each version of the product; there is an opportunity to review risks related to costs and compliance with the schedule; the customer gets used to the new technology gradually. Disadvantages: the functional system must be fully defined at the beginning of the iteration allocation life cycle; with constant changes, the structure of the system may be disturbed;

– Spiral model – using this model, the customer and the development team analyze the risks of the project and execute it in iterations. The next stage is based on the previous one, and at the end of each round - a cycle of iterations - a decision is made on whether to continue the project. Advantages: special attention is paid to risk management; additional features may be added at later stages; the flexible design is possible. Disadvantages: risk assessment at each stage is quite expensive; constant feedback and reaction of the customer can provoke new and new iterations, which can lead to a temporary delay in product development; more applicable for large projects;

– Iterative model – the process of creating software, which is carried out in small stages, during which the analysis of the obtained intermediate results is carried out, new requirements are put forward and previous stages of work are adjusted. Does not require the beginning of a complete specification of requirements. The creation begins with the implementation of a part of the functionality, which becomes the basis for determining further requirements. The product is created in such a way that, first of all, the basic functionality that works is designed. Then with each iteration, it is improved, and new features are added. Advantages: risk reduction – early detection of conflicts between requirements, models and project implementation; organization of effective feedback of the project team with the consumer, creation of a product that really meets his needs; rapid release of a minimally valuable product and the ability to bring the product to market and begin operation much earlier. Disadvantages: problems with the architecture and overhead costs - when working with chaotic requirements and without a developed global plan, the architecture of the program may suffer, and additional resources may be needed to bring it to an adequate appearance; there is no fixed budget and deadlines, and strong involvement of the customer in the process is required.

Authors Petersen K., Wohlin C., and Baca D. in the article [3] performed an analytical review of rigid methodologies and the problems that arise when using them. Rigid methodologies are often used in projects where late-stage changes are too expensive or impossible. For example, to create complex engineering structures (aviation, construction, etc.). Rigorous methodologies are also used to develop software in systems for military or medical needs. Also, an overview of the advantages and disadvantages of the Waterfall model is covered by the authors Adetokunbo A.A. Adenowo, and Basirat A. Adenowo in the publication [4].

Among the flexible methodologies, the following are known today:

– Lean – the concept of «Lean production», which is based on the optimization of the company's work, where the processes are focused on the final value (the developed project or the sold product) and the removal of those teams that do not create additional value. According to the conclusions of Grynko T., the advantages of this concept include the following: saving time, which will allow the completion of a larger number of projects; flexibility; involvement

of each team member in the optimization process; maximum focus on the consumer. Among the shortcomings, the author includes the need for extremely high qualifications of project managers or other management; does not contribute to increasing the scientific and technological level (R&D); needs a perfect information support system (ISS) and others [5];

– eXtreme Programming (XP) – a software development methodology that differs from other flexible methodologies in that it is used only in the field of software development. It cannot be used in another business. XP Principles: Simplicity, Communication, Feedback, Courage and Respect. Advantages: extreme programming; the customer receives exactly the product he needs; the code always works due to constant testing and continuous integration; pair programming; low risks. Disadvantages: it is difficult to predict the time spent on the project; the methodology works only with senior specialists; not suitable for large projects;

– Rational Unified Process (RUP) – a methodology that involves product development in the following stages: initial; specification; construction; implementation. Each of them includes one or more iterations;

– Dynamic Systems Development Model (DSDM) – a methodology that demonstrates a set of principles, defined types of roles and techniques. The principles are aimed at the main goal - to deliver the finished project on time and within the budget, with the ability to adjust requirements during development;

– Rapid Application Development (RAD) – the methodology of rapid development of applications, which involves the use of tools for visual modelling (prototyping) and development. RAD involves small development teams, deadlines of up to 4 months, and active involvement of the customer from the early stages. This methodology is based on the requirements, but there is also the possibility of changes during the development of the system. This approach allows you to reduce costs and reduce development time to a minimum;

– Extreme Programming (XP) – the methodology focused on constantly changing product requirements offers 12 approaches to achieve effective results in similar conditions. Among them: a quick plan and its constant change; simple architecture design; frequent testing; simultaneous participation of two developers in one task or even at one workplace; continuous integration and frequent small releases;

– Agile model – it is a powerful methodology for software development, which represents certain systems that determine the order of tasks, evaluation and control methods. Since 2000, in the practice of project management, the agile model has been widely used. Since then, methods of its use (Scrum & Kanban, XP, etc.) and software for Agile project management have been developed and implemented. This methodology is so progressive that it is also used in other areas (marketing, education, finance, construction, and others). Advantages: quick decision-making due to constant communications; risk minimization; easier work with documentation. Disadvantages: a large number of meetings and conversations, which can increase the time of product development; it is difficult to plan processes because the requirements are constantly changing; rarely used to implement large projects.

Many researchers are engaged in the study and research of project management methods. For example, the authors Brych V., and Peryt I., believe that in times of global and constant changes, the most effective is the use of flexible methodologies [6].

According to the researchers Abrahamsson P., Salo O., Ronkainen J., and Warsta J. agile model makes it possible to adjust the schedule of the project, make certain changes to the project during its implementation, constant improvement and flexible reactions to changes in requirements, potential and understanding the problems that need to be solved, etc [7].

In articles [8, 9], the authors conducted an analysis of the most popular software development methodologies, from which it follows that flexible methodologies better meet business needs and are the most popular among IT companies.

Therefore, the analysis of the methodologies used today in the information technology (IT) market showed that they are chosen based on the direction of the project, its budget, implementation terms and the ability to adapt to flexible business needs. At the same time, due to the lack of clear regulation of actions, the development of projects within the cycle is ensured in different ways. Therefore, the purpose of this article is to perform a comparative analysis of the use of flexible methodologies in the process of the development life cycle and their selection for the creation of a high-quality software product in clearly defined terms.

Performing an analysis of flexible methodologies and developing a block diagram for their selection

In order to achieve high project implementation results in any IT company, it is necessary to carefully plan each stage of its life cycle. For this, it is necessary to make a justified choice of a flexible methodology for each of the stages, which will make it possible to implement the project tasks. To do this, it is necessary to perform an in-depth analysis of the most used flexible Scrum and Kanban methodologies and develop an algorithm for their selection.

Comparing the Scrum and Kanban methodologies, the following criteria can be distinguished: in Scrum, all tasks are performed in sprints that last from 2 to 4 weeks, respectively. The following meetings are usually held: sprint planning, daily scrum, sprint review, and sprint retrospective. Usually, these meetings do not exceed 15 minutes. (table 1). In Kanban, meetings are optional. They can be of the following types: daily meeting, replenishment, deli-very planning meeting, service delivery meeting, operations review, risk review, strategy review.

The release occurs only when the product or some part of it is ready. Usually conducted several times a day or once a week.

Table 1

Scrum events

Events	Inspection	Adaptation
Sprint Planning	– Product Backlog; – Commitments Retrospective; – Definition of Done	– Sprint Goal; – Forecast; – Sprint Backlog
Daily Scrum	– Progress toward Sprint Goal	– Sprint Backlog; – Daily Plan
Sprint Review	– Product Increment; – Product Backlog (Release); – Market-business conditions	– Product Backlog
Sprint Retrospective	– Team & collaboration; – Technology & engineering; – Definition of Done	– Actionable improvements

In Scrum, a backlog is created with clear tasks to be completed in a sprint. Kanban has a workflow where all the tasks that must be completed are displayed and they are moved from one status to another. At the same time, there is no clear time frame planning.

There can be no changes or additions in a sprint in Scrum. At the beginning of the sprint, it is decided how many and which tasks can be done (sprint scope). Such changes are possible in Kanban. When a task is completed, a new task can be created.

Scrum meetings are usually attended by the product owner, scrum team, and scrum master. In Kanban, project management can be performed by a manager, there is no focus on the distribution of roles, attention is focused on the production of the project. The roles, artefacts, and practices in Scrum are listed in table 2.

Table 2

The roles, artefacts, and practices in Scrum

Roles (Who?)	Artefacts (What?)	Practices (How?)
Product Owner	– Product Backlog	– Sprint
Scrum team	– Sprint Backlog	– Sprint Planning Meeting
Scrum master	– Potentially Shippable Product	– Daily Standup
	– Sprint Burndown Chart	– Sprint Review
		– Sprint Retrospective

In Scrum, task boards are created for each new sprint. The organization of the board has the following parts: «To do», «In progress», «In testing» and «Done».

In Kanban, the board is one and the same. This is an advantage if the team is small because you can see on one board all the tasks, their relationships with each other and the progress of execution. Usually consists of columns «To do», «In progress», and «Done».

Work productivity in Scrum is measured in story points, that is, in the speed of completing tasks per sprint. Assessment of tasks is always available.

In Kanban, productivity is measured in the speed at which a task moves from the «To do» status to the «Done» status. Tasks are not always evaluated. A comparative analysis of Scrum and Kanban methodologies is given in table 3.

Table 3

Comparative analysis of Scrum and Kanban methodologies

№	Parameters	SCRUM	KANBAN
1	Visualization of the life cycle	+	+
2	Availability of iterations	+	+
3	Presence of backlog	+	-
4	Daily sprints	+	+
5	Big projects	-	+
6	Medium projects with the possibility of making changes	+	+
7	People and interaction come first	+	-
8	Processes and tools come first	-	+
9	Readiness for change	+	+
10	Development speed	+	+
11	Minimization of risks	+	+
12	Repeated sprints of fixed duration	+	-
13	Continuous process	-	+
14	Release at the end of each sprint after manager approval	+	-
15	The flow continues without interruption or at the discretion of the team	-	+
16	Roles: product owner, scrum master, scrum team	+	-
17	Roles: a team led by a manager	-	+
18	The main indicator is the speed of the team	+	-
19	The main indicator is time	-	+
20	During the sprint, changes are undesirable	+	-
21	Changes can happen at any moment	-	+
22	Self-organization of the team regarding the distribution of tasks	+	+

From table 3, it can be concluded that each of the methodologies has its own differences.

Table 3 can be used to create a block diagram of a Scrum or Kanban methodology selection program for a company's work organizations. However, for this, it is necessary to make several transformations with the table 3.

We perform the first transformation by discarding parameters that do not affect the selection. Namely, parameters by numbers: 1, 2, 4, 6, 9, 10, 11, 22. Because the parameters of the methodologies have the same positive value. The result of the first transformation is the reduction of the number of parameters to 14 (table 4).

Table 4

Parameters that affect the choice of methodology

№	Parameters	SCRUM	KANBAN
1	Presence of backlog	+	-
2	Big projects	-	+
3	People and interaction come first	+	-
4	Processes and tools come first	-	+
5	Repeated sprints of fixed duration	+	-
6	Continuous process	-	+
7	Release at the end of each sprint after manager approval	+	-
8	The flow continues without interruption or at the discretion of the team	-	+
9	Roles: product owner, scrum master, scrum team	+	-
10	Roles: a team led by a manager	-	+
11	The main indicator is the speed of the team	+	-
12	The main indicator is time	-	+
13	During the sprint, changes are undesirable	+	-
14	Changes can happen at any moment	-	+

The second transformation is the grouping of parameters according to the priority of selection. Parameter 2 (is the size of the project large) should be the first because a positive answer to it makes the choice of Kanban methodology. Because further choices do not make sense. The second parameter should be 1 (about the presence of a backlog), since a positive answer to it also makes further selection impractical. In addition, it is necessary to reformulate the parameters into questions.

The third transformation is a combination of parameters that are mutually opposite. In parameters 3 and 4, the user must make a choice that for him, people or tools are more important in the organization of the process. It is also necessary to combine parameters 5 and 6 (continuous process, or sprints of fixed duration); 7 and 8 (release at the end of each sprint after approval by the manager, or the flow continues without interruption or at the discretion of the team); 9 and 10; 11 and 12; 13 and 14. After combining the parameters, it is necessary to reformulate the combined parameters into questions in the same way as in the previous transformation.

The fourth transformation is replacing the symbol «-» with «no» and «+» with «yes».

Only after the last transformation, table 4 with parameters takes on a form that can be used to develop a block diagram of the program for choosing the company's work organization methodology (table 5).

Table 5

Data for the development of a block diagram of the methodology selection program

№	Parameters	SCRUM	KANBAN
1	Your project is big	**no**	**yes – end**
2	You will have a list of work tasks arranged in order of importance (backlog)	**yes – end**	**no**
3	In the first place you are:	people and their interaction **(yes)**	tools come first **(no)**
4	The process is:	repeated sprints of fixed duration **(yes)**	continuous teamwork **(no)**
5	The flow is:	release at the end of each sprint after manager approval **(yes)**	continues without interruption or at the discretion of the team **(no)**
6	Roles:	Product Owner, Scrum Master, Scrum Team **(yes)**	a team led by a manager **(no)**
7	The main indicator:	team speed **(yes)**	time **(no)**
8	Changes:	during the sprint, changes are undesirable **(yes)**	changes can happen at any moment **(no)**

As we can see from table 5, the first two parameters allow continuing the choice of methodology, if the answer is «no». The last six do not allow you to switch to the next one, because the program will end when you select any of the six listed items. However, there is a possibility of the so-called evaluation selection of the listed parameters. After choosing one of the two possible answer options, the chosen one will be given a weight equal to «1», and the other – «0». The next step is to count the number of units in the methodology column. The methodology is chosen if

the number of positive answers is greater. However, it can be seen from table 5 that the number of positive answers can be equal to three. In this case, you need to choose a mixed methodology for the organization of the company's work.

In the block diagram (fig. 1), the parameters from table 5 are given by numbers.

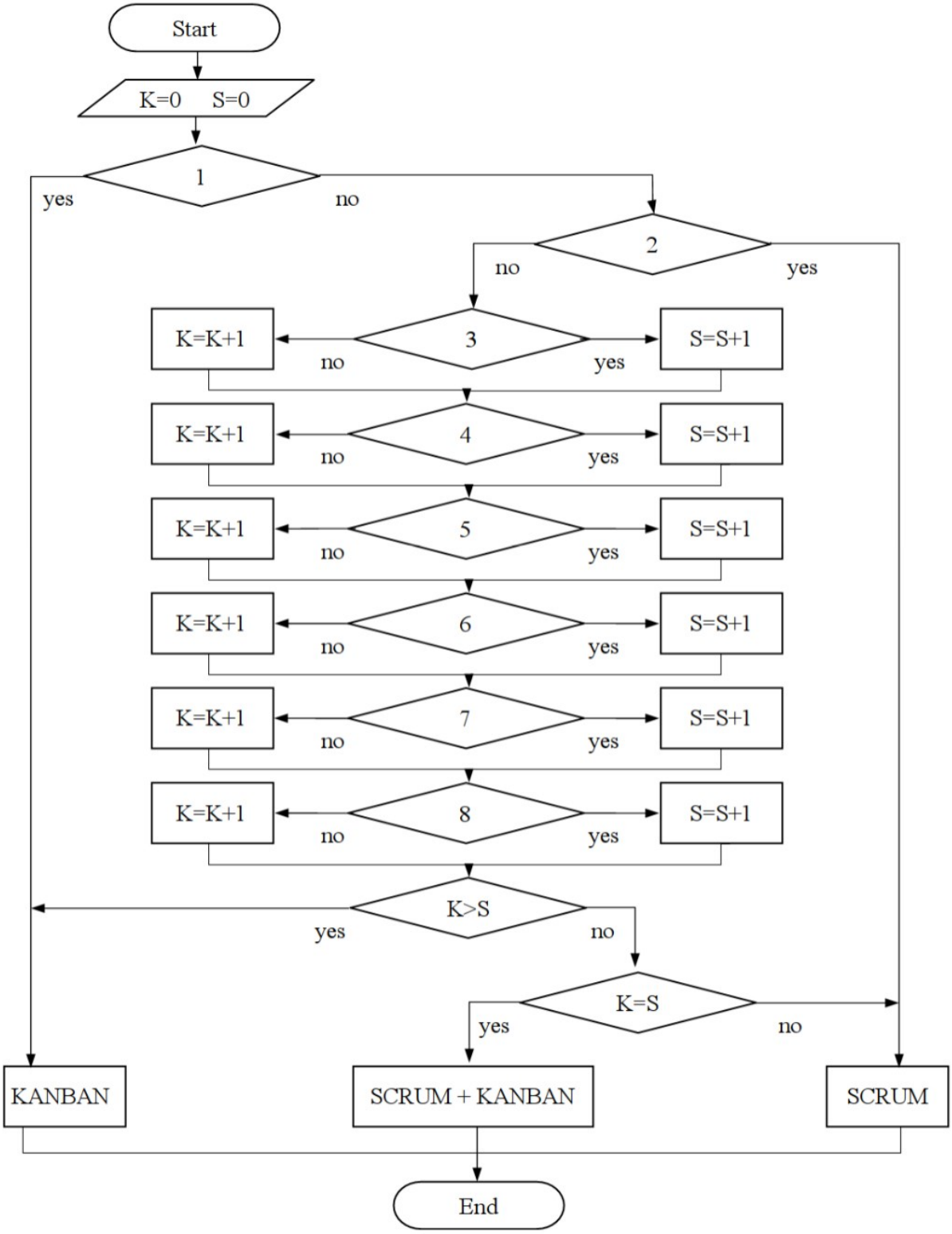


Fig. 1. Block diagram of Scrum or Kanban methodology selection program

Conclusions

Therefore, the analysis of the methodologies has shown that today there are a sufficient number of them, among which there are rigid and flexible ones. It was determined that most researchers prove in their works that

flexible methodologies are more widespread nowadays and are better adapted to the conditions of rapid changes in projects.

The article provides a comparative analysis of the most popular today's flexible Scrum and Kanban methodologies. A block diagram has been developed for an informed choice of Scrum or Kanban methodology. The article is based on research on the international and Ukrainian markets of IT companies.

The direction of the authors' future research is to conduct a more detailed analysis and comparison of rigid and flexible methodologies with the aim of using combined forms of project management.

References

1. Scrum Alliance. URL: https://www.scrumalliance.org/.
2. Continuous Improvement Through Enterprise Agility. URL: https://xscion.com/agile/.
3. Petersen K., Wohlin C., Baca D. (2009). The Waterfall Model in Large-Scale Development. In: Bomarius, F., Oivo, M., Jaring, P., Abrahamsson, P. (eds) Product-Focused Software Process Improvement. PROFES 2009. Lecture Notes in Business Information Processing, vol 32. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-02152-7_29.
4. Adetokunbo A.A. Adenowo, Basirat A. Adenowo. Software Engineering Methodologies: A Review of the Waterfall Model and Object-Oriented Approach. International Journal of Scientific & Engineering Research, Volume 4, Issue 7, July-2013.
5. Grynko T. Advantages and disadvantages of the «Lean production». URL: http://www.confcontact.com/2017-ekonomika-i-menedzhment/4_grinko.htm.
6. Brych V. Basics of business management of domestic economic entities / V. Brych, I. Peryt // Scholarly notes of the «KROK» University. Chapter 6. Management and marketing. - No3 (55). - 2019. - Pp. 82-93.
7. P. Abrahamsson, O. Salo, J. Ronkainen, J. Warsta. Agile Software Development Methods: Review and Analysis. URL: https://arxiv.org/abs/1709.08439.
8. Vavilenkova A. Analysis of flexible methodologies of software development for implementation in team projects / A. Vavilenkova // Visnyk NTU «KhPI». - №1 (7). – 2021. – Pp. 39-46.
9. Korzh B. Flexible models of teamwork management of engineering projects / B. Korzh // Bulletin of Taras Shevchenko National University of Kyiv. Economics, 2019; 6(207): Pp. 21-27.

Iryna Zasornova Ірина Засорнова	PhD, Associate Professor of Computer Engineering & Information Systems Department, Khmenlnytskyi National University https://orcid.org/0000-0001-6655-5023 e-mail: izasornova@gmail.com	Доцент кафедри комп'ютерної інженерії та інформаційних систем, Хмельницький національний університет
Sergii Lysenko Сергій Лисенко	DrSc (Engineering), Professor of Computer Engineering & Information Systems Department, Khmenlnytskyi National University https://orcid.org/0000-0001-7243-8747 e-mail: sirogyk@ukr.net	Доктор технічних наук, професор кафедри комп'ютерної інженерії та інформаційних систем, Хмельницький національний університет
Oleksandr Zasornov Олександр Засорнов	PhD, Associate Professor of Computer Engineering & Information Systems Department, Khmenlnytskyi National University https://orcid.org/0000-0001-9931-4154 e-mail: azasornov@gmail.com	Доцент кафедри комп'ютерної інженерії та інформаційних систем, Хмельницький національний університет

UDC 004.912
https://doi.org/10.31891/csit-2022-4-2

OLESIA BARKOVSKA
Kharkiv National University of Radio Electronics

PERFORMANCE STUDY OF THE TEXT ANALYSIS MODULE IN THE PROPOSED MODEL OF AUTOMATIC SPEAKER'S SPEECH ANNOTATION

The global spread and use of remote and online learning systems at various educational levels puts forward a number of requirements for existing systems and needs for expansion of functionality. The current problem in Ukraine is the unstable operation of the energy infrastructure due to frequent hostile shelling, so it is problematic for residents of Ukraine to join online classes on time, to listen to lectures by lecturers and teachers completely, to take part in conferences and master classes in full. This determines the need to provide the opportunity of familiarization with educational materials at a convenient time in a form convenient for understanding and mastering. The lecture recording provides access to audio files that are intended for listening, but are not intended for printed reproduction. Therefore, the expansion of existing digital educational platforms with the possibility of forming an annotation (summary, abstract) of a lecture and presenting it in the form of text-and-graphic materials for further use by course students on paper media is an urgent task and can improve the quality assessment of a remote educational resource from the point of view of the content and methodological aspect. The aim of the study is to create a generalized hybrid model of automatic annotation of the speaker's speech, which provides for the possibility of recognizing the speech, transforming the available data into text and, at the last stage, summarizing the given text, keeping only the important meaningful part of a lecture. The desired aim was achieved due to the creation of a generalized hybrid model of automatic annotation of input audio data, taking into account the effectiveness and features of existing methods of automatic text annotation obtained after converting speech into text. The uniqueness of this study is the use of marker words at the stage of text summarization, as well as the comparison of the efficiency of data processing at different stages of operation of this model when using different hardware. The results of computational experiments on graphics processing units with the Turing architecture showed that when the scope of input data increases by almost 30 times, the time also increases proportionally, but the use of a more powerful graphics processing unit NVIDIA Tesla T4 gives an speedup of more than 2.5 times compared to the graphics processing unit NVIDIA GeForce GTX GPU 1650 Mobile for both English and Ukrainian languages. For texts in the Ukrainian language, the text compression obtained (the ratio of the word count of the input text array to the word count in the resulting annotation) is 89.7%, for English – 94.15%. The proposed use of marker words showed an increase in the logical connection of input information internally, but obliges speakers to use predefined marker words to preserve the structure of the annotation formed.

Keywords: annotation, text, input data, language, abstracting, calculation, graphics processing unit, summarization.

ОЛЕСЯ БАРКОВСЬКА
Харківський національний університет радіоелектроніки

ДОСЛІДЖЕННЯ РОБОТИ МОДУЛЮ АНАЛІЗУ ТЕКСТУ У ЗАПРОПОНОВАНІЙ МОДЕЛІ АВТОМАТИЧНОГО АНОТУВАННЯ ПРОМОВИ СПІКЕРА

Глобальне поширення та використання систем дистанційного та он-лайн навчання на різних освітніх рівнях висуває ряд вимог до існуючих систем та потребує розширення функціоналу. Проблемою сьогодення в Україні є нестабільна робота енергетичної інфраструктури через часті ворожі обстріли, тому, приєднуватися до онлайн занять вчасно, слухати повноцінні лекції лекторів та учителів, приймати участь у конференціях та майстер-класах у повному обсязі, жителям України є проблематичним. Це обумовлює необхідність забезпечити можливість ознайомлення із навчальними матеріалами у зручний час узручному для розуміння та засвоєння вигляді. Запис лекції забезпечує доступ до звукових файлів, які припускаються прослуховування, але не призначені для друкованого відтворення. Тому, розширення існуючих цифрових освітніх платформ можливістю формування анотації (резюме, реферату) лекції та подання її у вигляді текстографічних матеріалів для подальшого використання слухачами курсу на паперових носіях, є завданням актуальним та здатне підвищити оцінку якості дистанційного освітнього ресурсу з погляду змістовно-методологічного аспекту. Метою дослідження є створення узагальненої гібридної моделі автоматичного анотування промови спікера, яка надає можливість розпізнавання мовлення, перетворення наявних даних в текст і останнім етапом проведення сумаризації даного тексту, зберігаючи лише важливу змістовну частину лекції. Поставлену мету було досягнуто завдяки створенню узагальненої гібридної моделі автоматичного анотування вхідних аудіо даних, враховуючи ефективність та особливості існуючих методів автоматичного анотування тексту, отриманого після конвертації промови у текст. Новизною даного дослідження є використання слів маркерів на етапі сумаризації тексту, а також порівняння ефективності обробки даних на різних етапах роботи даної моделі при використанні різного апаратного забезпечення. Результати обчислювальних експериментів на графічних процесорах із архітектурою Turing показали, що при збільшенні обсягів вхідних даних майже у 30 разів, час також збільшується пропорційно, але використання більш потужного графічного процесора NVIDIA Tesla T4 дає прискорення більше ніж у 2.5 рази порівняно із графічним процесором NVIDIA GeForce GTX 1650 Mobile як для англійської, так і для української мови. Для текстів українською мовою отримане стиснення тексту (відношення кількості сліх вхідного текстового масиву до кількості слів в отриманій анотації) становить 89,7%, для англійської мови – 94,15%. Запропоноване використання слів-маркерів показало підвищення логічного зв'язку вхідної інформації між собою, але зобовязує спікерів використовувати попередньо визначені слова-маркери для збереженн структури сформованої анотації.

Ключові слова: анотування, текст, вхідні данні, мова, реферування, обчислення, графічний процесор, сумаризація

Introduction

Information presented in text form is a valuable source of knowledge; however, it often needs to be effectively processed to get as much benefit as possible. Every year, the issue of creating an annotation (summary,

abstract) becomes more and more relevant [1, 2, 3]. For this purpose, it is necessary to compress text fragments to a shorter version, reduce the amount of the initial text while preserving key informational elements and content at the same time. Since it is a time-consuming and, as a rule, labor-intensive task to make annotation manually, the issue of automating this process is becoming increasingly popular in academic research.

An important component of the information space for remote education remains online lectures with experts, which can be held as a Q&A session and deal with questions from course students. Further access to online lecture materials should be convenient for understanding and mastering [4]. The lecture recording provides access to audio files that are intended for listening, but are not intended for printed reproduction.

Therefore, the expansion of existing digital educational platforms with the possibility of forming an annotation (summary, abstract) of a lecture and presenting it in the form of text-and-graphic materials for further use by course students on paper media is an urgent task, since it can improve the quality of presentation of educational information and the conditions of working therewith, as well as improve the quality assessment of a remote educational resource from the point of view of the content and methodological aspect [5].

That is why one of the areas of research is speech processing and conversion of audio files into text material, while keeping only important and relevant information. The key challenges include topic determination, interpretation, abstract generation, and its quality assessment. The most important tasks include identifying key phrases and using them to select sentences that will be included in the annotated text.

Text abstracting is the task of compressing a text fragment into a shorter version, reducing the amount of the original text while preserving key informational elements and content at the same time. Since manual text summarization is a time-consuming and, as a rule, labor-intensive task, the issue of automating the task is becoming increasingly popular and therefore is a strong motivation for academic research [6, 7].

There are important text summarization tasks related to NLP, such as classification of texts, answering to questions, summarization of legal texts, summarization of news and generation of headings. In addition, summarization can be integrated into these systems as an intermediate stage that contributes to reducing the length of a document [8].

In the age of big data, there has been an explosion in the amount of textual data from various sources. This text length is an invaluable source of information and knowledge that should be effectively summarized to be useful. The growing availability of documents requires comprehensive research in the domain of natural language processing for automatic text summarization. Figure 1 shows a diagram of a typical text summarization workflow.

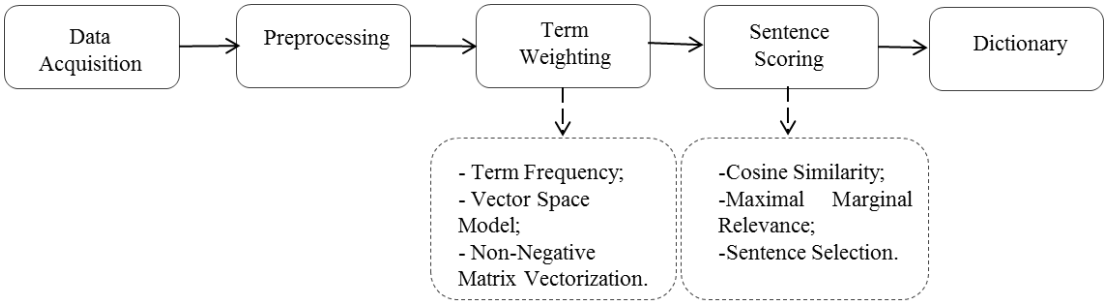


Fig. 1. Text summarization workflow

Most of existing approaches to text summarization model the problem as a classification problem that decides whether to include a sentence in the summary or not. Other approaches have used information on the topic, latent semantic analysis (LSA), sequence-to-sequence models, reinforcement learning and adversarial processes.

Related works. Research task rationale

The relevance of the work described above consists in increasing the efficiency and relevance of forming an annotation of a speaker's (lecturer's, expert's, teacher's, etc.) speech. The analysis of existing methods of annotating text data also proves the existing interest in NLP methods and in particular in methods of text summarization when performing academic research. The analysis of the problem area has shown that there are two general approaches to automatic abstracting (Table 1):

extraction (extractive approach) [9]. When extracting, the content is extracted from the input data, but the extracted content is not changed in any way. The methods of this approach characterize the existence of a function of evaluation of the importance of information block. As a rule, the importance of a sentence is determined by the importance of the words therein;

abstraction (abstractive approach) [10.]. Abstractive methods build an internal semantic presentation of the original text, and then use this representation to create an abstract. It involves the generation of new words and phrases that do not appear in the input text to report the most useful information from the original text.

Table 1

Generalization of analysis of the features of extractive and abstractive approaches to abstracting

	Extractive method of abstracting	**Abstractive method of abstracting**
Advantages	More simple than the abstractive approach, since it is based on copying pieces of the input text based on the determination of key phrases; it is easier to ensure basic levels of grammar and accuracy.	provides for application of additional knowledge due to the use of deep learning; the resulting abstract is closer to an abstract that can be generated by a human, since it uses a semantic analysis of the entire input text.
Disadvantages	does not provide for paraphrasing, inclusion of additional knowledge for high-quality summarization.	requires deep knowledge of the developer in the domain of artificial intelligence and computer linguistics.

At present, there are some high-tech solutions from different companies, but each of them has its advantages and disadvantages as discussed below (Table 2).

Dragon Anywhere is voice recognition software. This solution allows the user for dictating large documents without limitation on the time of dictation or numbers of pages. If a mistake is made during dictation, there is an option to correct it or edit the previous sentence using simple voice commands, such as “correct”. The correction menu that appears will provide a contextual list of alternative phrases to choose from.

Table 2

Overview of existing solutions in the domain of STT and annotation of text arrays

	Dragon Anywhere	**Amazon Transcribe**	**QuillBot**
Advantages	high accuracy of voice recognition (~ 99%); no word count limit; several ways to exchange documents.	high accuracy of voice recognition; possibility of interaction with other solutions of the Amazone ecosystem.	there is an option to add a browser extension; ease of use.
Disadvantages	lack of text summarization option; possibly cutthroat prices; it may take time to learn the built-in commands.	high cost of use; lack of the text summarizing option (there is an option of using separate modules, which will lead to an increase in the cost of use); an understanding of the AWS ecosystem is required.	works only with English language; lack of ability to dictate the text; has limitations when using the free version.

Amazon Transcribe is an automatic speech recognition service that makes it easy to add speech-to-text options to any application. Transcribe functions allow for obtaining audio, creating and reviewing easy-to-read transcripts, improving accuracy with customization and filtering content to ensure customer privacy.

QuillBot is a paraphrasing and summarizing tool that helps millions of students and professionals to reduce their time of writing by more than half by using the most advanced AI to rewrite any sentence, paragraph or article. It has both free and premium version. There is also access to use the API.

Aims and tasks of the work

The aim of the study is to create a generalized hybrid model of automatic annotation of the speaker’s speech, which provides for the possibility of recognizing the speech, transforming the available data into text and, at the last stage, summarizing the given text, keeping only the important meaningful part of a lecture.

Since the reliability of information contained in the educational resources of remote courses is one of the key requirements for digital educational platforms, cutting down the emergence of false or distorted data during the conversion of audio sequence into text data for further semantic analysis is the primary aim of the work.

The uniqueness of this study is the use of marker words at the stage of text summarization, as well as the comparison of the efficiency of data processing at different stages of operation of this model when using different hardware [11, 12, 13].

To achieve the desired aim, the following tasks should be solved:

- ✓ creation of a generalized hybrid model of automatic annotation of input audio data;
- ✓ review and analysis of existing methods of automatic annotation;
- ✓ adaptation of input text annotation methods for different computing architectures;
- ✓ evaluation of the timing of operation of the text analysis module;
- ✓ analysis of the results obtained.

Results and Discussion

The paper proposes a generalized hybrid model of automatic annotation of input audio data (Figure 2).

The automatic speech recognition (ASR) module accepts input of sound recording in WAV format, cleans the audio sequence using a deep neural network, and converts the cleaned audio sequence into text [14, 15].

The text analysis module accepts input of the deliverables from the speech recognition (ASR) module in the form of a JSON object. Text filtering takes place at the stage of transition of the JSON object from the ASR module

to the text summarization module. Next, the selection of key characteristics of the text and the extraction of the most significant fragments of the text using the mT5 model (pre-trained multilingual transformer for 101 languages), which is an extension of the Text-to-Text Transfer Transformer (T5) model.

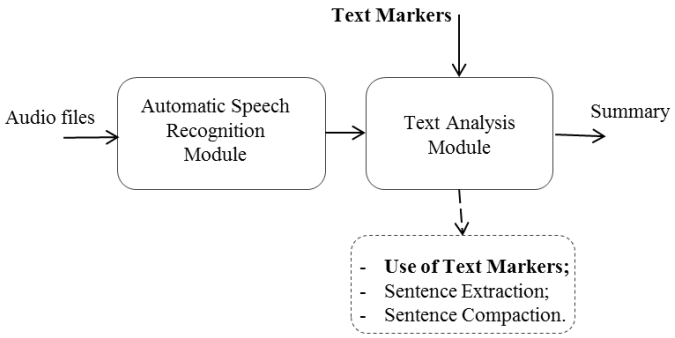


Fig.2 . Generalized hybrid model of automatic speaker's speech annotation

This solution was trained in 101 languages on a corpus of Common Crawl web pages, and supplemented with the XL-SUM dataset (covering 45 languages, highly abstract, concise and high-end as evidenced by human and internal evaluation). The data in different languages was sampled so that the balance between rare and popular web page languages could be adjusted.

When presenting the experiments, the results of summarizing texts for the Ukrainian and English languages were studied. Please find the results of benchmarking the evaluation of XL-SUM test sets according to the ROUGE metric in Table 3.

Table 3

Benchmarking of XL-SUM test sets

Language	ROUGE-1	ROUGE-2	ROUGE-3
English	37.601	15.1536	29.8817
Ukrainian	23.9908	10.1431	20.9199

The paper proposes an idea for creating a service for filtering the input text using marker words as described below.

The primary idea of the Text markers sub-module is to break the input text into fragments. That is, a separate json file is created with the "chopped" text between two word markers. This functionality creates an option of abstracting the separate fragments of the text without mixing unrelated information, which can lead to the loss of the sense of information. After the abstracting stage, the data received is "glued" into one document of the following type: "Word marker: content". According to the study conducted, this improvement is aimed at increasing the logical connection of input information internally.

Please find the scheme of performance of this solution in Figure 3. The sequence of stages of the text analysis module with modification by adding text markers is as follows: text and text markers are input, the text is parsed and cut into "pieces", then for each "piece" separately the summarization process is launched, and the process finishes with the stage of gluing the data into a single document.

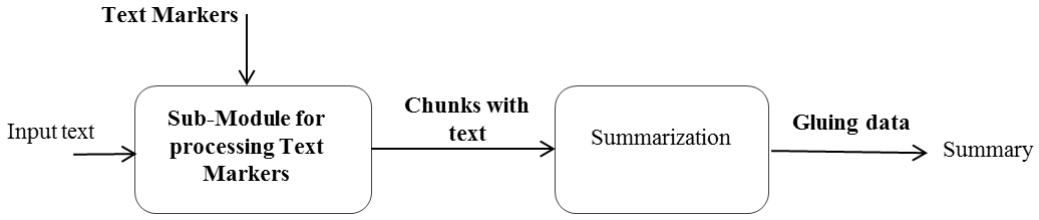


Fig. 3. Scheme of modification of the text analysis module due to the use of text markers

The performance evaluation of the text analysis module demonstrated quite high results for the task of abstracting texts in the Ukrainian and English languages. This approach to summarization uses the abstractive method using parallel computing.

Please find the results of comparison of the input and output text in Table 4. The text obtained from the ASR module is used for comparison.

Table 4

Comparison of text compression for summarization completed

Language	Initial word count	Final characters count	Text compression, %
Ukrainian	136	14	89.7
English	188	11	94.15

Please find below the examples taken for comparison of the performance of the summarization module in the table above.

In general, it is necessary to point out the quite high quality of text abstracting even in Ukrainian. It should be noted that study in the NLP domain for text summarization has been conducted for a relatively long time for many languages, but the leader in terms of high rates is English.

Computational experiments have been conducted with the use of computers with different performance. The following hardware has been used as an available estimator on a personal computer – central processor Intel Core i7-9750H (2.6-4.5 GHz), graphics processing unit NVIDIA GeForce GTX 1650 Mobile. The characteristics of the hardware on the remote cloud solution are as follows – central processor Intel Xeon 2.30GHz, graphics processing unit NVIDIA Tesla T4.

Please find the time spent by the text analysis module for processing the input data in the Ukrainian language in Table 5.

Table 5

Comparison of the time spent on summarization of the text in Ukrainian

Word count	Time spent on a personal computer (Ukrainian), sec	Time spent on a cloud solution (Ukrainian), sec
25	0.4	0.37
136	3.2	0.92
725	13.6	5.13

According to the results shown in the diagram (Figure 4), there is a trend to increase in the data processing time with the growth of the text dictionary, which is the expected result. It is possible to see a time reduction in data processing for a more powerful graphics card, namely NVIDIA Tesla T4, as compared to NVIDIA GeForce GTX 1650 Mobile PC graphics card.

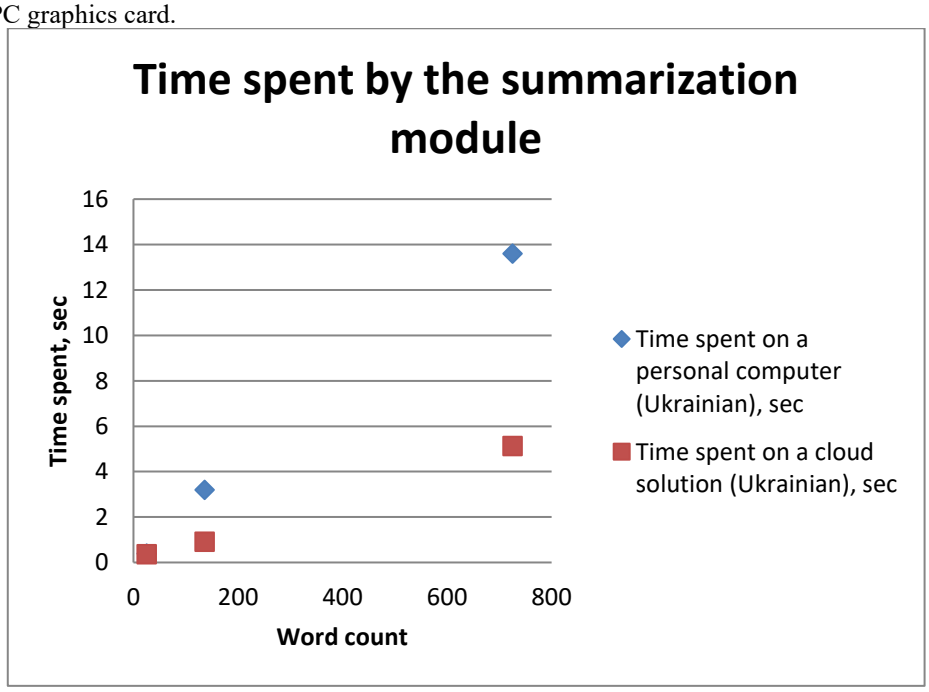


Fig. 4. Diagram of time spent for processing text in Ukrainian

Please find the time spent by this module for processing input data in English in Table 6.

Table 6

Comparison of the time spent on summarizing the text in English

Word count	Time spent on a personal computer (English), sec	Time spent on a cloud solution (English), sec
32	0.38	0.34
187	3.3	0.94
713	12.7	4.98

According to the results obtained, a diagram was built demonstrating the time reduction in data processing for the Ukrainian language. In the same way as in the case above, using a more powerful graphics card can show a significant increase in data processing for larger text content. Please find the diagram in Figure 5.

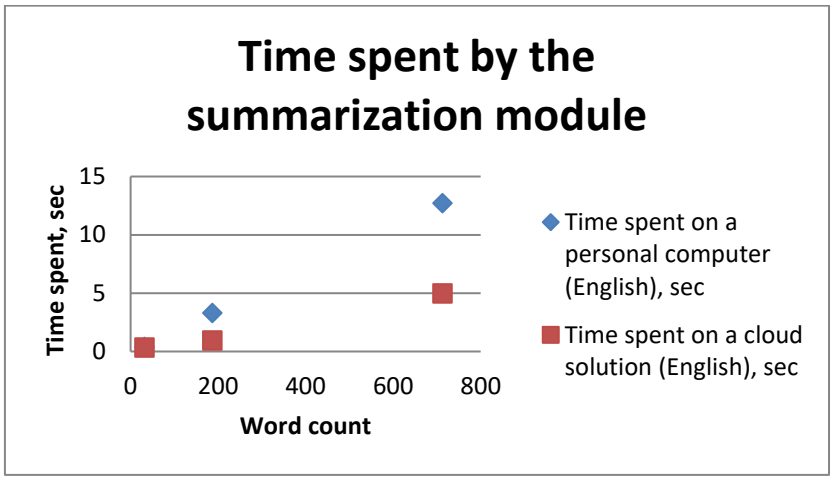


Fig. 5. Time consumption diagram for processing text in English

Therefore, we can make a conclusion on the effectiveness of speeding up data processing in this module with a more powerful video card. For the Ukrainian language, there is a significant speedup with a larger scope of input data, namely if we take into account the test results of 136 and 725 input words, the average speedup will be about 34%. Based on the results for 725 input words, the more there are input words, the higher is the speedup from a more powerful graphics card.

As to the processing of text in English, the result is slightly faster and the speedup is observed as well for a more powerful video card. The average speedup result for the input of 187 and 713 is 34% as well.

The studies conducted to improve the performance of the text analysis module due to the use of text markers proves that the solution developed compensates for the problem of the loss of context of a document and additionally with the use of parallel computing, does not critically load the system due to the distribution of independent annotation for selected text "pieces" that have been separated by the user with pre-determined text markers. However, the proposed approach creates a limitation for the speaker, namely it compels the speaker to use marker words.

Conclusions

The expansion of existing digital educational platforms with the possibility of forming an annotation (summary, abstract) of a lecture and presenting it in the form of text-and-graphic materials for further use by course students on paper media is an urgent task and can improve the quality assessment of a remote educational resource from the point of view of the content and methodological aspect. The aim of the study is to create a generalized hybrid model of automatic annotation of the speaker's speech, which provides for the possibility of recognizing the speech, transforming the available data into text and, at the last stage, summarizing the given text, keeping only the important meaningful part of a lecture. The desired aim was achieved due to the creation of a generalized hybrid model of automatic annotation of input audio data, taking into account the effectiveness and features of existing methods of automatic text annotation obtained after converting speech into text. The uniqueness of this study is the use of marker words at the stage of text summarization, as well as the comparison of the efficiency of data processing at different stages of operation of this model when using different hardware. The results of computational experiments on graphics processing units with the Turing architecture showed that when the scope of input data increases by almost 30 times, the time also increases proportionally, but the use of a more powerful graphics processing unit NVIDIA Tesla T4 gives an speedup of more than 2.5 times compared to the graphics processing unit NVIDIA GeForce GTX GPU 1650 Mobile for both English and Ukrainian languages. For texts in the Ukrainian language, the text compression obtained (the ratio of the word count of the input text array to the word count in the resulting annotation) is 89.7%, for English – 94.15%. The proposed use of marker words showed an increase in the logical connection of input information internally, but obliges speakers to use predefined marker words to preserve the structure of the annotation formed.

Further research and improvement of the proposed generalized model of automatic annotation of the speaker's speech is the possibility of deploying this model in cloud solutions such as Amazon Web Services or Google Cloud Platform to prevent data loss in the event of war or natural disasters. Cloud solutions ensure the reliability of data storage and processing due to the creation of data snapshots and replication thereof between servers that have different geographical locations. And additionally in the event of unavailability of the necessary hardware – the use of dedicated capacities of cloud solutions.

References

1. N. Bharti, S. N. Hashmi and V. M. Manikandan, "An Approach for Audio/Text Summary Generation from Webinars/Online Meetings," *2021 13th International Conference on Computational Intelligence and Communication Networks (CICN)*, 2021, pp. 6-10, doi: 10.1109/CICN51697.2021.9574684.
2. M. Kirmani and A. K. Shukla, "Systematic review of methods used in text summarization," *2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*, 2022, pp. 1048-1052, doi: 10.1109/ICACITE53722.2022.9823831.
3. Zhou, M., Duan, N., Liu, S., & Shum, H. Y., "Progress in neural NLP: modeling, learning, and reasoning*", Engineering*, 6(3), 275-290, https://doi.org/10.1016/j.eng.2019.12.014
4. Barkovska, Olesia, Viktor Khomych, and Oleksandr Nastenko. "Research of the text processing methods in organization of electronic storages of information objects," *Innovative Technologies and Scientific Solutions for Industries*, 1(19), (2022), https://doi.org/10.30837/ITSSI.2022.19.005
5. Barkovska, O., Pyvovarova, D., Kholiev, V., Ivashchenko, H., & Rosinskiy, D., "Information Object Storage Model with Accelerated Text Processing Methods", *In COLINS,* 2021 (April), pp. 286-299.
6. Ramezani, Majid, and Mohammad-Reza Feizi-Derakhshi. "Automated text summarization: An overview." *Applied Artificial Intelligence,*2014, 28.2, pp. 178-215.
7. Rahul, S. Adhikari and Monika, "NLP based Machine Learning Approaches for Text Summarization," *2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC),* 2020, pp. 535-538, doi: 10.1109/ICCMC48092.2020.ICCMC-00099.
8. El-Kassas, W. S., Salama, C. R., Rafea, A. A., & Mohamed, H. K. (2021). "Automatic text summarization: A comprehensive survey", *Expert Systems with Applications*, 165p.
9. R. Mishra, V. K. Panchal and P. Kumar, "Extractive Text Summarization - An effective approach to extract information from Text," *2019 International Conference on contemporary Computing and Informatics (IC3I)*, 2019, pp. 252-255, doi: 10.1109/IC3I46837.2019.9055636.
10. D. Singhal, K. Khatter, T. A and J. R, "Abstractive Summarization of Meeting Conversations," *2020 IEEE International Conference for Innovation in Technology (INOCON)*, 2020, pp. 1-4, doi: 10.1109/INOCON50539.2020.9298305
11. Olesia Barkovska, Oleg Mikhal , Daria Pyvovarova , Oleksii Liashenko , Vladyslav Diachenko and Maxim Volk, "Local Concurrency in Text Block Search Tasks",. *International Journal of Emerging Trends in Engineering Research*, 8(3), March 2020, pp.690-694, DOI:10.30534/ijeter/2020/13832020
12. Barkovska O., Pyvovarova D. and Serdechnyi V., "Pryskorenyj alghorytm poshuku sliv-obraziv u teksti z adaptyvnoju dekompozycijeju vykhidnykh danykh". [Accelerated word-image search algorithm in text with adaptive decomposition of input data], *Systemy upravlinnja, navighaciji ta zv'jazku,* 4 (56), 2019, 28-34. (in Ukrainian)
13. O. Barkovska, P. Rusnak, V. Tkachov and T. Muzyka, "Impact of Stemming on Efficiency of Messages Likelihood Definition in Telegram Newsfeeds," *2022 IEEE 3rd KhPI Week on Advanced Technology (KhPIWeek),* 2022, pp. 1-5, doi: 10.1109/KhPIWeek57572.2022.9916415.
14. Alshemali, B., & Kalita, J., "Improving the reliability of deep neural networks in NLP: A review", *Knowledge-Based Systems*, 191, 105210.
15. I. Mykhailichenko, H. Ivashchenko, O. Barkovska and O. Liashenko, "Application of Deep Neural Network for Real-Time Voice Command Recognition," *2022 IEEE 3rd KhPI Week on Advanced Technology (KhPIWeek),* 2022, pp. 1-4, doi: 10.1109/KhPIWeek57572.2022.9916473.

Olesia Barkovska **Олеся Барковська**	Ph.D., Associate Professor of Department of Electronic Computers, Kharkiv National University of Radio Electronics, Kharkiv, Ukraine https://orcid.org/0000-0001-7496-4353 e-mail: olesia.barkovska@nure.ua	Доцент к.т.н., доцент кафедри електронних обчислювальних машин, Харківський національний університет радіоелектроніки, Харків, Україна

UDC 004.9: 347.151
https://doi.org/10.31891/csit-2022-4-3

TETIANA HOVORUSHCHENKO, SERGII ALEKSOV,
SNIZHANA TALAPCHUK, OLEKSII SHPYLYUK, VIKTOR MAGDIN
Khmelnytskyi National University

OVERVIEW OF THE METHODS AND TOOLS FOR SITUATION IDENTIFICATION AND DECISION-MAKING SUPPORT IN THE CYBERPHYSICAL SYSTEM «SMART HOUSE»

The technology of a smart house is mostly understood as a system that combines a number of subsystems that provide comfortable living conditions for residents in the room and make it possible to significantly reduce energy costs. A house is called smart if it has a certain computer or control system for managing engineering equipment. "Smart House" should be designed so that all services can be integrated with each other with minimal costs (in terms of finances, time and effort), and their maintenance would be organized in an optimal way.

The "Smart House" system should competently allocate resources, reduce operating costs and provide a clear control and management interface. Such an intelligent system should be able to recognize specific planned and emergency situations occurring in the home and respond to them (make decisions) in accordance with the given program. Therefore, today the urgent task is to recognize the situation and support decision-making in the "Smart House" cyber-physical system.

The conducted overview of methods and tools for situation identification and decision-making support in the "Smart House" cyber-physical system showed that: in existing solutions, situation recognition occurs only for one of the groups of the system of managed housing functions or does not occur at all; existing solutions provide decision-making support for only one of the groups of the system of managed housing functions or do not provide it at all; the available solutions do not provide for the possibility of assessing the sufficiency of information for decision-making in the "Smart House" cyber-physical system.

Therefore, there is a need to develop such methods and tools for situation identification and decision-making support in the "Smart House" cyber-physical system, which would: perform situation recognition for all 5 groups of the system of managed housing functions; provide the decision-making support for all 5 groups of the system of managed housing functions; perform an assessment of the sufficiency of information for making all decisions in the "Smart House" cyber-physical system, which will be the focus of the authors' further efforts.

Keywords: cyber-physical system "Smart House", system of managed housing functions, housing microclimate management; housing lighting control; housing security system; management of multimedia systems of housing; control of household appliances and power grid of housing.

ТЕТЯНА ГОВОРУЩЕНКО, СЕРГІЙ АЛЕКСОВ,
СНІЖАНА ТАЛАПЧУК, ОЛЕКСІЙ ШПИЛЮК, ВІКТОР МАГДІН
Хмельницький національний університет

ОГЛЯД МЕТОДІВ І ЗАСОБІВ РОЗПІЗНАВАННЯ СИТУАЦІЇ ТА ПІДТРИМКИ ПРИЙНЯТТЯ РІШЕНЬ У КІБЕРФІЗИЧНІЙ СИСТЕМІ «РОЗУМНИЙ БУДИНОК»

Під технологією розумного будинку здебільшого розуміють систему, що об'єднує в собі ряд підсистем, які забезпечують комфортні умови проживання мешканців у приміщенні та надають змогу суттєво зменшити витрати енергоносіїв. Будинок називається розумним, якщо в ньому наявна певна комп'ютерна чи контролююча система управління інженерним оснащенням. «Розумний будинок» повинен бути спроектований так, щоб всі сервіси могли інтегруватися один з одним з мінімальними витратами (з точки зору фінансів, часу і трудомісткості), а їх обслуговування було б організовано оптимальним чином.

Система «Розумний будинок» повинна грамотно розподіляти ресурси, знижувати експлуатаційні витрати і забезпечувати зрозумілий інтерфейс контролю і управління. Така інтелектуальна система повинна вміти розпізнавати конкретні заплановані та надзвичайні ситуації, що відбуваються у помешканні, і реагувати на них (приймати рішення) відповідно до заданої програми. Отже, на сьогодні актуальним завданням є розпізнавання ситуації та підтримки прийняття рішень у кіберфізичній системі «Розумний будинок».

Проведений огляд методів і засобів розпізнавання ситуації та підтримки прийняття рішень у кіберфізичній системі «Розумний будинок» показав, що: в наявних рішеннях розпізнавання ситуацій відбувається лише для однієї з груп системи керованих функцій житла або не відбувається взагалі; наявні рішення передбачають підтримку прийняття рішень лише для однієї з груп системи керованих функцій житла або не передбачають її взагалі; в наявних рішеннях не передбачається можливість оцінювання достатності інформації для прийняття рішень у кіберфізичній системі «Розумний будинок».

Отже, виникає необхідність в розробленні таких методів і засобів розпізнавання ситуації та підтримки прийняття рішень у кіберфізичній системі «Розумний будинок», які б: виконували розпізнавання ситуацій для всіх 5 груп системи керованих функцій житла; передбачали підтримку прийняття рішень для всіх 5 груп системи керованих функцій житла; виконували оцінювання достатності інформації для прийняття всіх рішень у кіберфізичній системі «Розумний будинок», на що й будуть спрямовані подальші зусилля авторів.

Ключові слова: кіберфізична система «Розумний будинок», система керованих функцій житла, керування мікрокліматом житла; керування освітленням; система безпеки; керування системами мультимедіа; керування побутовою технікою та електромережею.

Introduction

"Smart House" is a living environment of a modern type, organized for people to live with the help of automation and high-tech devices that form an intelligent control system to ensure the coordinated and automatic operation of all engineering networks of the house [1]. The technology of a smart house is mostly understood as a system that combines a number of subsystems that provide comfortable living conditions for residents in the room and make it possible to significantly reduce energy costs [2]. A house is called smart if it has a certain computer or control system for managing engineering equipment [2]. "Smart House" should be designed so that all services can be integrated with each other with minimal costs (in terms of finances, time and effort), and their maintenance would be organized in an optimal way [3].

The "Smart House" system competently allocates resources, reduces operating costs and provides a clear control and management interface. Such an intelligent system should be able to recognize specific planned and emergency situations occurring in the home and respond to them according to a given program: one of the systems, according to the programmed algorithm, can control the behavior of others [1].

An important feature and property of the "Smart House", which distinguishes it from other ways of organizing the living environment, is that it is the most progressive concept of human interaction with the living space, when the resident of the house chooses one of the programmed scenarios, and the automated control system in accordance with external and internal conditions sets the parameters and monitors the operating modes of all engineering systems and electrical devices [1].

The system of managed housing functions consists of *five main groups* [1]:

1) housing microclimate management;
2) housing lighting control;
3) housing security system;
4) management of multimedia systems of housing;
5) control of household appliances and power grid of housing.

Creating and maintaining an optimal *home microclimate* is the most important condition for high efficiency, productive rest and health of residents of a house or apartment. The climate control system in the room makes it possible to set the optimal level of temperature, humidity, the amount of fresh air inflow, control the operation of the air filtration system, and create an individual climate system for each family member, in particular, for a child [1].

Lighting control of the residential environment is divided into control of three types of lighting: natural, artificial and light dynamics (Fig. 1). In order to regulate the illumination of the premises by natural daylight and shade the windows in the evening, the "Smart House" system controls the positions of the blinds and shutters, as well as the mechanical opening and closing of the curtains. The intelligent system for managing artificial lighting sources regulates the brightness and number of lighting devices for each individual room or functional zone, depending on the time of day, weather conditions, and the type of activity of the residents at a specific time. One of the important possibilities of the "smart house" is the creation of dynamic light scenarios, when pressing one button turns on the optimal lighting for a particular situation [1, 4, 5].

Fig. 1. Typical implementation of a lighting control system

The security system in the "Smart House" system has several areas of protection: protection against intrusion, protection against water and gas leaks, fire safety, video surveillance system, alarm buttons and simulation of the presence of the owners at home [1, 5, 6].

In addition to the service function, the "Smart House" is also equipped with *internal multimedia systems* for the entertainment of the owners of the house and their guests: multiroom (multi-zone audio and video distribution system), media server, home theater (Fig. 2) [1].

Fig. 2. Typical implementation of the "multiroom" system

The management of household appliances and the electrical network is an important part of the overall complex of intelligent management of the housing environment. The following components can be attributed to it: scenarios for switching on or off the equipment, control of individual sockets or their groups, control of household appliances (Fig. 3) [1, 7].

Fig. 3. Typical implementation of remote control in the house

Systems of intelligent control of the housing environment have a wide range of functional purposes, perform numerous operations according to many scenarios [1, 8].

A homeowner doesn't need to have deep programming knowledge to operate such a powerful system, as all scenarios are pre-programmed and configured to suit the needs of the family. It is enough for home owners to control the functions of the "Smart House" through control devices with an intuitive interface [1, 9, 10].

Therefore, today *the urgent task* is to recognize the situation and support decision-making in the "Smart House" cyber-physical system.

Overview of the methods and tools for situation identification and decision-making support in the cyberphysical system "Smart House"

Let's conduct an overview of known methods and tools for situation identification and decision-making support in the "Smart House" cyber-physical system, highlighting their advantages and disadvantages.

In the paper [2], fuzzy logic algorithms are used to determine the comfortable conditions of stay in the "Smart House" system, in particular, to calculate the comfortable temperatures. A basic term set is formed for each linguistic variable. For example, for the variables "temperature inside the room", "air temperature of the atmosphere", such a set consists of four terms: "cold", "neutral", "warm", "hot". After the selection of linguistic variables, term sets are formed and membership functions are constructed, production rules for the model are compiled. This technique makes it possible to determine how to adjust the temperature to comfortable values by evaluating the internal and external air temperature using the rules of fuzzy logic.

The study [11] developed the rolling-horizon optimization model with a recurrent neural network-driven predicting, which is developed for interactively prediction of uncertainty and optimization of battery energy storage operations in residential smart houses in an iterative fashion. The proposed model can be used for optimizing battery energy storage operations in residential smart houses and for efficiently utilizing solar power.

Home energy management systems are used for management of energy consumption in smart houses. The research [12] presented home energy management strategy (OHEM-algorithm) based on the improved binary particle swarm optimization, which intended for optimization of customer satisfaction and electric cost, for getting the accurate, optimal, and desirable solutions for power consumption in the smart homes, for lower the cost of electricity and the user's conformity.

The paper [13] proposes a Smart Apartment Building model, in which multiple distributed power sources are shared by multiple consumers for reducing the operation costs and carbon emissions through the implementation of highly efficient operation methods.

The paper [14] proved that the fuzzy logic with Multi-class Support Vector Machine (SVM) method, which is realized as the fuzzy trapezoidal membership function for each sample within the hyper-sphere and as a linear function of the selected sample's distance in the non-linear SVM hyperplane, is effective in selection of the rules to make decision to the control in temperature and humidity.

Paper [15] made the OTP-based door opening system using Arduino and GSM, which generates the one-time password on mobile phone for unlocking the door and is much safer than the traditional key-based system.

Paper [16] develops the Internet-of-Things-based indoor, comfortable, environmental, and real-time monitoring system for the smart house, which consists of the temperature-and humidity-sensing module and the lightness module. In this system, improved particle swarm optimization (IPSO) is used for creating the ideal and comfortable environment.

The paper [17] investigated the appliance of electrical use as a means for detecting the presence/absence of residents (for example, people suffering from dementia, elderly people living alone, home quarantine) with using the several machine learning algorithms.

The research [18] is devoted to the full state feedback and feed forward control method for determination of the best control theory for control of the servo motor in the smart window systems, which is ised for improving the air circulation and for better automation of the air circulation.

The paper [19] presents the Internet-of-Things-based smart kitchen system, which automatically detects the temperature, monitors the humidity level, includes built-in gas detection sensors for detection of gas leaks in the kitchen, provides the remotely control of the appliances (ovens, freezers, and air conditioners) using the mobile phone. This system is realized on an Arduino board with the Internet connection. The system’s goal is remotely control devices (switches, fans, and lights) by any Android smartphone.

The paper [20] presents an Emergy-based methodological approach for assessing the effectiveness of integration of the IoT-based sensing systems into smart buildings for reducing their environmental impacts and energy consumption.

The research [21] focused on the addition of nodes into the IoT-based smart home infrastructure, on the design, implementation and testing the hardware and software of the ESP-Mesh-based smart home system (using the ESP8266) with 3 different nodes – mechanical (door lock), temperature & humidity sensors, electrical (fan, generic power switch, or power plug).

Paper [22] proposed the new data driven method for accurate indirect heat accounting in apartment buildings, which provided the measurements or estimations of the difference of temperature between the indoor environment and the heat transfer fluid, because of which the heating bill's error is reduced by 20%–50%.

The paper [23] proposes a development method and TOPPERS Embedded-Component System on the basis of the embedded components for devices for improving the development efficiency of smart homes' electrical equipment, increasing the electrical equipment's scalability and reducing the developmental complexity.

The paper [24] investigated to the development of the voice-activated home automation system, which integrates the Artificial Intelligence, Internet of Things, Natural Language Processing, Blockchain for a cost-effective and efficient interacting with household equipment.

The paper [25] proposed the method of optimal energy consumption in the smart houses on the basis of the optimal scheduling the household appliances, considering demand side management and techno-economic indices in electrical grids.

The paper [26] is devoted to the development of Internt-of-Things-based system for control heating and cooling within the residential housing, which accurately identifies whether it should be cooled or heated, so that energy is not wasted.

The paper [27] proposed Smart Exterior Home Management System for automatically managing the house's exterior activities without the human efforts (automating the water motor, notifications of the house members about receeiving the posts or deliveries, car parking shed and gate, ring a calling bell, if any person is detected near the main door of the house).

The paper [28] proposed the secure user authentication and key agreement scheme using physical unclonable functions for preventing the security problems, used Real-or-Random model and Burrows-Abadi-Needham logic for verification of the session key security and mutual authentication, used too the Automated Validation of Internet Security Protocols and Applications tool for simulation of the scheme resistance to security attacks.

The paper [29] proposed an automatic control heating and domestic hot water system into a single-family house with installing sensors, PID regulators and actuators, with monitoring control system in the SIEMENS TIA Portal software tool via intelligent interface. This approach increases energy efficiency and reduces the energy costs in the building.

Paper [30] expanded a Secure Smart Home Automation System using Arduino UNO and Wi-Fi technology using Face recognition gadgets with the purpose of the implementation of greater protection to the users and greater effectiveness of the software tool, greater luxury and greater usefulness for the old humans or handicapped.

The paper [31] aims to develop software, which is capable of controlling all electrical devices of a house based on a Raspberry-based control system with the smartphone tools for ensuring the adequate securities.

Results & Discussion

The conducted overeview of methods and tools for situation identification and decision-making support in the "Smart House" cyber-physical system showed that:

1) in existing solutions, recognition of situations occurs only for one of the groups of the system of managed housing functions or does not occur at all;

2) existing solutions provide decision-making support for only one of the groups of the system of managed housing functions or do not provide for it at all;

3) the existing solutions do not provide for the possibility of assessing the sufficiency of information for decision-making in the "Smart House" cyber-physical system.

So, based on the critical analysis of methods and tools for situation identification and decision-making support in the cyber-physical system "Smart House", during which the above-mentioned shortcomings were highlighted, there is a need to develop such methods and tools for situation identification and decision-msking support in the cyber-physical system "Smart house", which would: perform recognition of situations for all 5 groups of the system of managed housing functions; provide the decision-making support for all 5 groups of the system of managed housing functions; evaluate the sufficiency of information for making all decisions in the "Smart House" cyber-physical system (Fig. 4).

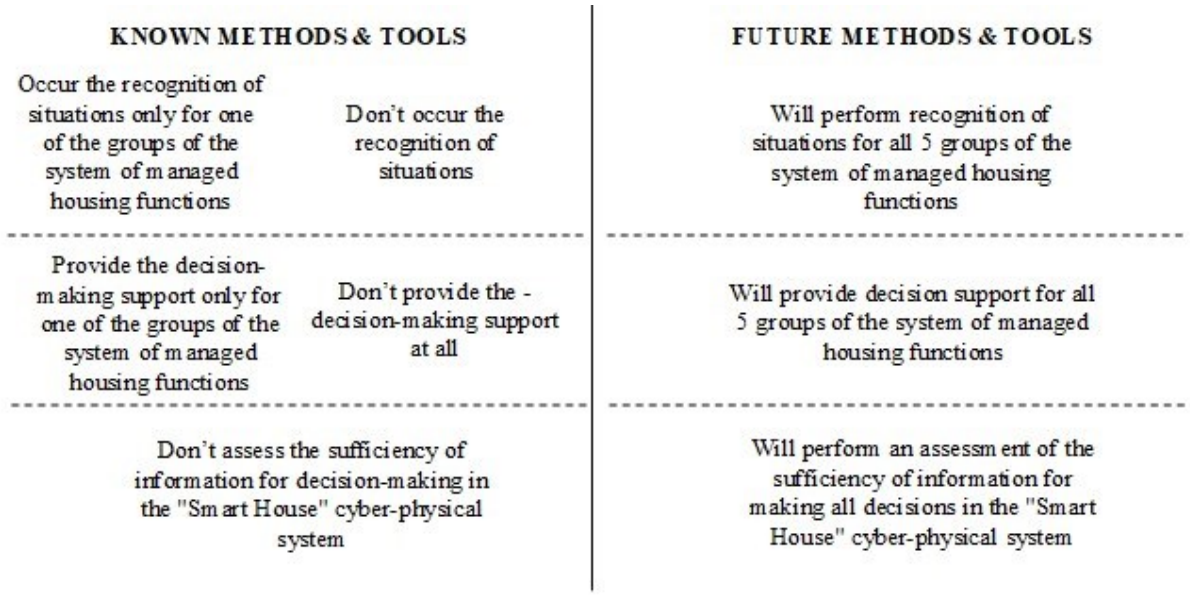


Fig. 4. The role of the proposed approach in the cyber-physical system "Smart House"

Conclusions

The "Smart House" system should competently allocate resources, reduce operating costs and provide a clear control and management interface. Such an intelligent system should be able to recognize specific planned and emergency situations occurring in the home and respond to them (make decisions) in accordance with the given program. Therefore, today the urgent task is to recognize the situation and support decision-making in the "Smart House" cyber-physical system.

The conducted overview of methods and tools for situation identification and decision-making support in the "Smart House" cyber-physical system showed that: in existing solutions, situation recognition occurs only for one of the groups of the system of managed housing functions or does not occur at all; existing solutions provide decision-making support for only one of the groups of the system of managed housing functions or do not provide it at all; the available solutions do not provide for the possibility of assessing the sufficiency of information for decision-making in the "Smart House" cyber-physical system.

Therefore, there is a need to develop such methods and tools for situation identification and decision-making support in the "Smart House" cyber-physical system, which would: perform situation recognition for all 5 groups of the system of managed housing functions; provide the decision-making support for all 5 groups of the system of managed housing functions; perform an assessment of the sufficiency of information for making all decisions in the "Smart House" cyber-physical system, which will be the focus of the authors' further efforts.

References

1. M. Maslova. "Smart House": bibliographic index. URL: https://zounb.zp.ua/wp-content/uploads/2021/07/Rozumnij-budinok-pokazhchik-6.04.21-s-oblozhkoj.pdf.
2. I. Yurchak, P. Vyshynskyi. Application of Fuzzy Logic Algorithms in Smart Home Systems. Computer Systems and Networks. 2018. Vol. 905. Pp. 142-148.
3. S. Kukunin. Development of a holistic methodology for the organization of "Smart House" type systems within the framework of the "Internet of Things" paradigm. Computer-integrated technologies: education, science, production. 2020. Vol. 38. Pp. 40-45.
4. I. Shostak, M. Danova, O. Feoktystova. An approach to the robotization of the functioning processes of the "Smart House" system based on the Internet of Things. The XIII International Scientific and Practical Conference “Integrated intelligent robotic complexes”: Proceedings (Kyiv (Ukraine), May 19-20, 2020). Kyiv, 2020. Pp. 48-49.
5. V. Teslyuk, Kh. Beregovska, V. Beregovskyi. Model of operation of subsystems of lighting and protection of an intelligent building. Scientific bulletin of NLTU of Ukraine. 2013. Vol. 23. Issue 10. Pp. 297-303.
6. O. Boreiko, V. Teslyuk, O. Berezsky. Development of components of the "Intelligent House" video surveillance system based on Raspberry Pi. Modeling and information technologies. 2014. Vol. 71. Pp. 66-71.
7. I. Dontsov, O. Bezvesilna. Use of artificial intelligence in home automation and energy saving. The XI All-Ukrainian Scientific and Practical Conference of the Students and PhD Students “A look into the future of instrument building”: Proceedings (Kyiv (Ukraine), May 15-16, 2018). Kyiv, 2018. Pp. 505-508.
8. O. Polyakova. Classification of functional constituent elements of the system of intelligent management of the environment in housing design. Bulletin of the Kyiv National University of Technology and Design. Technical sciences. 2016. Vol. 4 (100). Pp. 133-140.
9. D. Fedorov. Increasing the comfort of life with the help of "Safe House" intelligent machines. Scientific notes of the Small Academy of Sciences of Ukraine. Pedagogical sciences. 2018. Vol. 12. Pp. 179-185.
10. I. Sribna, A. Aleksandrov. Interactive automatic system "Smart House". Communication. 2019. Vol. 3. Pp. 55-58.
11. S. Abedi, S. Kwon. Rolling-horizon optimization integrated with recurrent neural network-driven forecasting for residential battery energy storage operations. International Journal of Electrical Power and Energy Systems. 2023. Volume 145. Article number 108589.
12. A. Mohammad, S. Ansari, F. Ali, I. Ashraf. Home Energy Management System with Improved Binary PSO. Lecture Notes in Electrical Engineering. 2023. Vol. 926. Pp. 873 – 881.
13. K. Tamashiro, E. Omine, N. Krishnan, A. Mikhaylov, A. M. Hemeida, T. Senjyu. Optimal components capacity based multi-objective optimization and optimal scheduling based MPC-optimization algorithm in smart apartment buildings. Energy and Buildings. 2023. Vol. 2781. Article number 112616.
14. K. Devi Thangavel, U. Seerengasamy, S. Palaniappan, R. Sekar. Prediction of factors for Controlling of Green House Farming with Fuzzy based multiclass Support Vector Machine. Alexandria Engineering Journal. 2023. Vol. 62. Pp. 279 – 289.
15. P. Srinivasan, R. Sabeenian, B. Thiyaneswaran, M. Swathi, G. Dineshkumar. OTP-Based Smart Door Opening System. Lecture Notes on Data Engineering and Communications Technologies. 2023. Vol. 131. Pp. 87 – 98.
16. W.-T. Sung, S.-J. Hsiao. Creating Smart House via IoT and Intelligent Computation. Intelligent Automation and Soft Computing. 2023. Vol. 35. Issue 1. Pp. 415 – 430.
17. A. Lentzas, D. Vrakas. Machine learning approaches for non-intrusive home absence detection based on appliance electrical use. Expert Systems with Applications. 2022. Vol. 21030. Article number 118454.
18. R. A. Lestari, U. Y. Oktiawati. Full state feedback and feed forward control of servo smart window using MATLAB/Simulink. Indonesian Journal of Electrical Engineering and Computer Science. 2022. Vol. 28. Issue 3. Pp. 1355 – 1362.
19. C. A. U. Hassan, J. Iqbal, M. Khan, S. Hussain, A. Akhunzada, M. Ali, A. Gani, M. Uddin, S. Ullah. Design and Implementation of Real-Time Kitchen Monitoring and Automation System Based on Internet of Things. Energies. 2022. Vol. 15. Issue 18. Article number 6778.
20. T. Kumar, R. Srinivasan, M. Mani. An Emergy-based Approach to Evaluate the Effectiveness of Integrating IoT-based Sensing Systems into Smart Buildings. Sustainable Energy Technologies and Assessments. 2022. Vol. 52. Article number 102225.
21. S. Fuada, Hendriyana. UPISmartHome V.2.0 – A Consumer Product of Smart Home System with an ESP8266 as the Basis. Journal of Communications. 2022. Vol. 17. Issue 7. Pp. 541 – 552.
22. Y. Stauffer, F. Saba, R. Carrillo, M. Boegli, A. Malengo, A. Hutter A. Smart sensors network for accurate indirect heat accounting in apartment buildings. Journal of Building Engineering. 2022. Vol. 461. Article number 103534.
23. J. Y. Jiang, F. Qi, H. Oyama, H. Nagashima, T. Azumi. ECHONET Lite Framework Based on Embedded Component Systems. ECTI Transactions on Computer and Information Technology. 2022. Vol. 16. Issue 1. Pp. 74-83.
24. S. Ansar, K. Jaiswal, S. Aggarwal, S. Shukla, J. Yadav, N. Soni. Smart Home Personal Assistants: Fueled by Natural Language Processor and Blockchain Technology. The 2022 2nd International Conference on Interdisciplinary Cyber Physical Systems: Proceedings (Chennai, May 9-10, 2022). Chennai, 2022. Pp. 113-117.
25. I. Muda, N. Dwijendra, T. Awsi, B. Bashar, M. Majeed. Optimal Energy Scheduling of Appliances in Smart Buildings Based on Economic and Technical Indices. Environmental and Climate Technologies. 2022. Vol. 26. Issue 1. Pp. 561 – 573.

26. W. Yaici, E. Entchev, M. Longo. Internet of Things (IoT)-Based System for Smart Home Heating and Cooling Control. The 2022 IEEE International Conference on Environment and Electrical Engineering and 2022 IEEE Industrial and Commercial Power Systems Europe: Proceedings (Prague, June 28 – July 1, 2022). Prague, 2022. Code 182192.

27. C. Prasad, Y. Srikanth, P. R. Rao, K. Sreedhar. Smart exterior home management system using Arduino Uno and Tinkercad. AIP Conference Proceedings. 2022. Vol. 2418. Article number 030039.

28. Y. Cho, J. Oh, D. Kwon, S. Son, J. Lee, Y. Park. A Secure and Anonymous User Authentication Scheme for IoT-Enabled Smart Home Environments Using PUF. IEEE Access. 2022. Vol. 10. Pp. 101330 – 101346.

29. K. Osman, M. Petic, T. Alajbeg, M. Stefic. Comparison of the theoretical mathematical model and the experimental approach in the development of an automatic control system in a smart family house. The 2022 7th International Conference on Smart and Sustainable Technologies: Proceedings (Split, July 5-8, 2022). Split, 2022. Code 182191.

30. A. Nirmala, V. Asha, P. Chandra, H. Priya, S. Raj. IoT based Secure Smart Home Automation System. The 2022 IEEE Delhi Section Conference: Proceedings (Online, February 11-13, 2022). Online, 2022. Code 178847.

31. Y. Tjandi, M. Paloboran, M. Yahya, A. Idkhan. Raspberry-based control system for the future house. Tehnicki Vjesnik. 2021. Vol. 28. Issue 6. Pp. 2115 – 2120.

Tetiana Hovorushchenko Тетяна Говорущенко	DrSc (Engineering), Professor, Head of Computer Engineering & Information Systems Department, Khmenlnytskyi National University https://orcid.org/0000-0002-7942-1857 e-mail: govorushchenko@gmail.com	Доктор технічних наук, професор, завідувач кафедри комп'ютерної інженерії та інформаційних систем, Хмельницький національний університет
Sergii Aleksov Сергій Алексов	PhD student of Computer Engineering & Information Systems Department, Khmenlnytskyi National University e-mail: aleksov@gmail.com	Аспірант кафедри комп'ютерної інженерії та інформаційних систем, Хмельницький національний університет
Snizhana Talapchuk Сніжана Талапчук	MSc student of Computer Engineering & Information Systems Department, Khmenlnytskyi National University e-mail: snizhanatalapchuk@gmail.com	Магістрантка кафедри комп'ютерної інженерії та інформаційних систем, Хмельницький національний університет
Oleksii Shpylyuk Олексій Шпилюк	Student of Computer Engineering & Information Systems Department, Khmenlnytskyi National University e-mail: oleksa280804@gmail.com	Студент кафедри комп'ютерної інженерії та інформаційних систем, Хмельницький національний університет
Viktor Magdin Віктор Магдін	Junior researcher of the Scientific and Research Department, Khmenlnytskyi National University e-mail: vityok555@gmail.com	Молодший науковий співробітник науково-дослідної частини, Хмельницький національний університет

UDC 004.855.5
https://doi.org/10.31891/csit-2022-4-4

TARAS RUDNYK, OLEG CHERTOV
National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute"

FORECASTING THE RESULTS OF THE PRESIDENTIAL ELECTIONS IN FRANCE BASED ON TWITTER DATA

This paper presents the study to collect, store and analyze data from Twitter to forecast French presidential election results, compared to sociological polls. The first and probably the most important step of the research is to collect, store and clean data, the whole result depends on the amount and quality of data. In the next step of research, datasets are analyzed. Lastly, complete report and visualizations are provided. In the study, we propose modern technics, mathematical algorithms, and machine learning approaches to analyze big amounts of data from the Twitter social network in order to forecast the 2022 French presidential election results. The determined outcome is compared with sociological polls and the real results of elections.

In the conducted research modern types of media are compared to select the best one for election prediction. Selected Twitter social network as the one with the most appropriate data and availability to download big amounts of useful information. The approach based on the usage of Python programming language, Selenium browser emulation and MongoDB database was used to collect, store and clean data about the main French election candidates – Emmanuel Macron and Marine Le Pen. The research was made from August 2021 until the election itself in April 2022. The determined outcome is compared with sociological polls and the results of elections and showed that analysis of social network data could be a good alternative to traditional sociological polls as it shows the same trends month by month and well predicted the win of Emmanuel Macron in elections. Moreover, the proposed approach has its benefits compared to sociological polls such as always being fresh, and close to real-time information, the price of research is much lower and could be reused for the next parliamentary or presidential elections with a small modification.

The research could be extended and adapted for other countries. Currently, the proposed algorithms and mathematical models showed good results in the French and Ukraine elections. It works well with English, French, Ukrainian and Russian languages. This allows us to claim that it will also work fine with other Latin or Cyrillic alphabets but for Asian or Arabic languages more research would be needed. Twitter is a good choice for European and American countries. In the future, other social networks should be considered for the countries in which it is not so popular.

Keywords: political rating, sociological poll, Twitter, Python, Selenium, data collection, machine learning, natural language processing.

ТАРАС РУДНИК, ОЛЕГ ЧЕРТОВ
Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського»

ПРОГНОЗУВАННЯ РЕЗУЛЬТАТІВ ВИБОРІВ У ФРАНЦІЇ НА ОСНОВІ ДАНИХ З TWITTER

У цій статті представлено дослідження збору, зберігання та аналізу даних із Twitter для прогнозування результатів президентських виборів у Франції у порівнянні з соціологічними опитуваннями. Першим і, мабуть, найважливішим кроком дослідження є збір, зберігання та очищення даних, оскільки весь результат залежить від кількості та якості даних. На наступному етапі дослідження проводиться аналіз наборів даних. В кінці надається повний звіт і візуалізація отриманих результатів. У дослідженні ми пропонуємо сучасну техніку, математичні алгоритми та підходи машинного навчання для аналізу великих обсягів даних із соціальної мережі Twitter, щоб спрогнозувати результати президентських виборів у Франції 2022 року. Отриманий результат порівнюється із даними соціологічних опитувань та фактичними результатами виборів.

У проведеному дослідженні порівнюються сучасні види медіа, щоб вибрати найкраще для прогнозування виборів. Вибрана соціальна мережа Twitter як така, що має найбільш відповідні дані та доступність для завантаження великої кількості корисної інформації. Підхід, заснований на використанні мови програмування Python, емуляції браузера Selenium і бази даних MongoDB, використовувався для збору, зберігання і очищення даних про головних кандидатів на виборах у Франції – Еммануеля Макрона і Марін Ле Пен. Дослідження проводилося з серпня 2021 року до самих виборів у квітні 2022 року. Визначений результат порівнюється з соціологічними опитуваннями та результатами виборів і показує, що аналіз даних соціальних мереж може бути хорошою альтернативою традиційним соціологічним опитуванням, оскільки він показує ті самі тенденції місяць за місяцем і добре передбачив перемогу Еммануеля Макрона на виборах. Більше того, запропонований підхід має свої переваги порівняно з соціологічними опитуваннями, такі як: завжди свіжа та наближена до реального часу інформація, ціна дослідження значно нижча та може бути повторно використана для наступних парламентських чи президентських виборів із невеликою модифікацією.

Дослідження можна розширити та адаптувати для інших країн. Наразі запропоновані алгоритми та математичні моделі показали хороші результати на виборах у Франції та Україні. Добре працюють з англійською, французькою, українською та російською мовами. Це дозволяє нам стверджувати, що вони також добре працюватимуть з іншими латинськими чи кириличними алфавітами, але для азіатських чи арабських мов потрібні додаткові дослідження. Twitter є хорошим вибором для країн Європи та Америки.

Ключові слова: політичний рейтинг, соціологічне опитування, Twitter, Python, Selenium, збір даних, машинне навчання, обробка природної мови.

Introduction

Nowadays there is a lot of data on the Internet. The modern world allows people to exchange opinions around the world about the different variety of topics. Some sources, like tv channels or newspapers, carefully prepare well-structured information and share it on social media. It is a job that is done by much fewer people than by Internet users. Almost everyone nowadays has an account on one or even all social networks such as Twitter, Facebook,

Instagram, or TikTok. Compared to social media in own pages, there are much more users. The information mostly is not well structured and could contain mistakes.

Everyone who analyzes social networks faces one or all listed below challenges:

- How to collect data?
- How to store big amounts of data?
- How to clean and transform data into a dataset?
- How to analyze data?
- How to visualize results?
- How to make a conclusion and create a report?

This research presents an approach that answers the listed above challenges that we faced while we were trying to forecast the results of the 2022 presidential elections in France based on data collected from Twitter. The elections in France were one of the hottest political topics at the beginning of the year 2022. Two candidates – Emmanuel Macron and Maria Le Pen were close to winning the election accordingly to social polls. Moreover, at some period time closer to the election date the rating of Maria Le Pen was growing, and at the same time rating of Emmanuel Macron was falling which was a pretty interesting situation to consider if we can achieve the same result as social polls using data collected from Twitter social network and predict the result of elections?

Related works

Data collection is a popular task in modern research. The lack of publicly available datasets motivates researchers to collect their own data which is often not an easy task. In [1] Yuji Roh, Geon Heo and Steven Euijong Whang conducted a survey on how data could be collected, cleaned, and labeled for machine learning. Accordingly, to the research, each step could have a modification, for example, labeling could be done not only by people but also using data programming or fact extraction. The gaps in data could be covered by generating synthetic data. From a machine learning perspective, results could be enhanced in different ways for example by improving the model or improving data. Before collection, it is important to understand which types of data there are on the chosen platform and which of them, we need. In [2] Hai Liang et al divide data into three types: content, behavior and network structure. For each type of data, authors suggest different approaches to harvest it, for example, content and behavior data – random selection, network structure – probability (or uniform) sampling, breadth-first search (BFS) sampling, and random walk (RW) sampling. For web, harvesting authors suggest using APIs or web scraping. The first type is also used in our research when it is possible but for scraping social networks traditional approach via HTML parsing is not working, therefore, browser emulation via Selenium web driver is used.

Nowadays the amount of textual data across the Internet is extremely large. The task of structuring and analyzing such data is impossible through manual human work. Much more practical to develop and use text analyzing and mining technics to automate these processes. Perhaps Noah Chomsky was the first linguist that started syntactic theories by introducing in 1957 syntactic structures. He defined a set of rules based on universal grammar. In 1965 [3] Noah Chomsky categorized syntactic theories into speech recognition (Higher Level) and natural language (Lower Level). Later in 1967 [4], Charles Hockett found some drawbacks in Noah Chomsky's study, the most important part was that in his study language is a well-defined, stable structure without mistakes which was possible only in rare, idealized conditions. Nowadays in an analysis of Internet sources such as social networks we can see that messages contain a lot of mistakes, aren't well structured and can be in different languages even with a mix of words from 2 or more languages. Text mining and natural language processing technics found their implication in many fields over the years. For example, in [5] S.-H. Liao, P.-H. Chu, and P.-Y. Hsiao showed that for decade from 2000 to 2011 years text mining techniques were applied to a variety of fields such as academics, industry, web applications and others.

There is a lot of research on social networks text analysis. For example, in [6] authors use modern natural language processing technics to identify fake news in social networks. The research described that usually text analysis starts from tokenization, punctuation, special characters removal, stopwords removal, spell checking, named entity recognition and stemming. In the next stage authors used vector representation of words because, for the calculation of machine learning models, data that could be operated mathematically was needed. TF-IDF [7] algorithm was used to analyze the texts further. Terms frequency (TF) – how many times this word appears in a document (sentence), Inverse Document Frequency (IDF) is the natural logarithm of the total number of documents divided by the total number of documents that contain this certain word + 1. All listed above technics are used in our research. For our dataset, we tried TF-IDF, Word2Vec [8] and Doc2Vec [9] technics and chose the best-performing one for us – Word2Vec.

In our [10] most recent study we conducted research on Ukrainian presidential elections, proposed several algorithms to determine the rating of politicians, detect dates when news affected ratings the most and identify specific news which influenced grows or fall of the rating. Experimental results conducted on Ukrainian President Volodymyr Zelenskyy page on Twitter show that the proposed approach allows not only to detect of ratings and their changes but detect news that influenced such changes the most. In the proposed research the model was enhanced with a new data collection algorithm and natural language processing technics.

Objectives: This study sets the complex problem of choosing data source, collecting information, storing it and analyzing it to detect the political rating of presidential election candidates. The choice of each of the listed above

steps can significantly improve or decrease the achieved results. It is important not only to calculate the rating at some point in time but also to have all updates as close as possible to real time and represent it in an understandable comfortable report to allow the user of the system to react quickly to each important event that caused these changes.

Data collection, preparation and analysis

Data sources. Media could be conditionally divided into 5 types:

1) printed media (newspapers, magazines);

2) broadcast media (radio, television);

3) outdoor media (billboards, posters);

4) websites (news sites, blogs);3) social networks: Twitter (https://twitter.com/), Facebook (https://www.facebook.com/), Instagram (https://www.instagram.com/), TikTok (https://www.tiktok.com/), Linkedin (https://www.linkedin.com/), Pinterest (https://www.pinterest.com/).

Most of the types are controlled by some company or person. Therefore, content is filtered and may contain paid articles or videos, which promote needed results. For our research, we need a place where people freely discuss their thoughts with coverage among ordinary people who share opinions not for money. For the listed above criteria social networks are the best match.

Nowadays there are a lot of social networks in the world. Each one of them potentially could be used for forecasting the results of the elections. Which social network to choose among Twitter, Facebook, Instagram, TikTok, Linkedin and Pinterest? Instagram, TikTok and Pinterest are mostly for photos or videos which are hard to analyze and hard to create, which means the amount of political content would be less than with text-based social networks. Linkedin doesn't match the research because of job orientation and low level of political posts. Both Twitter and Facebook are good choices for the research with a lot of users who discuss politics. Accordingly, to Statista (https://www.statista.com/statistics/284435/social-network-penetration-france/) Facebook is the most popular social network in France. The Facebook API is very limited, and it is almost impossible to download a big amount of data from this social network as it is very protected from data collection. Twitter API compared to Facebook's API allows to collect of much more data with less effort. With help of the official Twitter API, it is possible to collect tweets, the number of likes, retweets, and personal data such as user biography, age, location, website, and date joined. Taking to account all the conditions we decided to use Twitter as a data source.

Workflow for election results forecast. Figure 1 presents the general workflow for data collection, storing, processing, visualization, and reporting from the Twitter social network. Official Twitter API is limited, therefore, to collect data for the research hybrid approach is used. Twitter API is used as much as possible to collect data but the type of data that isn't accessible from API is collected by mimicking browser behavior using the Selenium web driver. As data storage MongoDB is used because of the following benefits:

- high availability of data with automatic fast data recovery. In our case it is important not to lose Tweets to have a whole historical dataset;
- in-build sharding solution. Conducting research over years in different countries requires a lot of space to store raw data. Sharding allows to separate of large databases into smaller, faster, more easily managed parts;
- unstable schema. In the beginning, we did not know what kind of data we would be able to download, and which parts would be useful for the research. With an unstable schema, it is easy to add or remove fields over time.

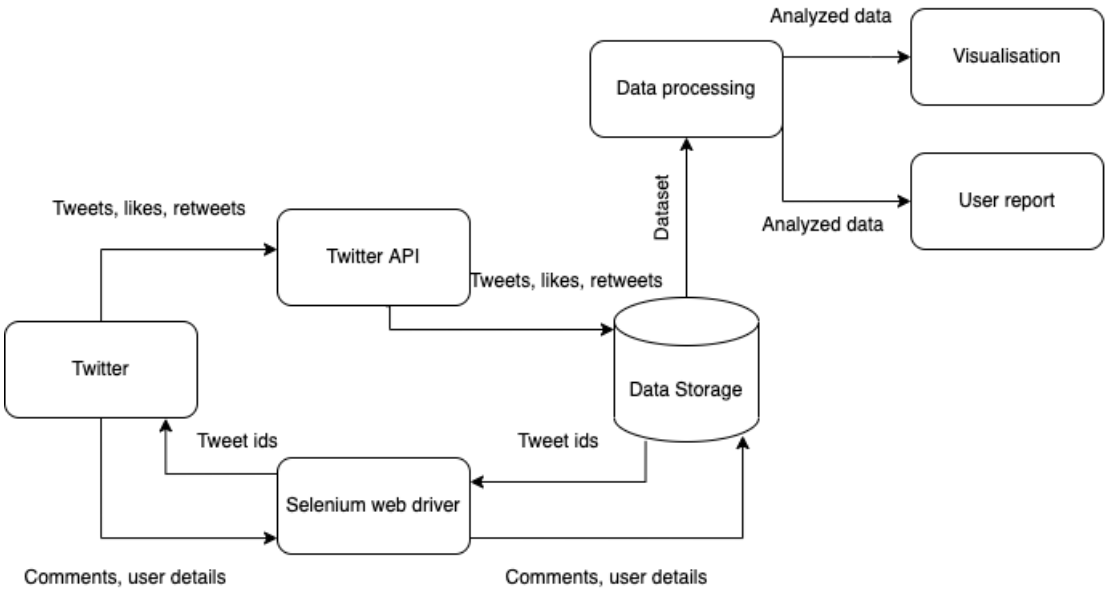


Fig. 1. General workflow for data collection, storing, processing, visualization, and reporting from Twitter social network

From MongoDB data storage CSV datasets are formed for data processing. For each politician, a separate dataset was used. After successful data processing visualization and user report are being built. The report contains whole statistics and comments while visualization – charts which may be used to quickly understand the current situation.

Figure 2 presents a detailed diagram of the data processing of the Twitter CSV dataset. In the research French elections are considered. Tweets about it are mostly in French and English languages. Therefore, other languages are omitted by the algorithm. Twitter text is hard to analyze as a whole. The dataset contains as small one-sentence messages as big messages containing multiple sentences. As the first step for each language text is split into sentences. The next steps perform the following natural language processing techniques:

- remove stop words;
- stemming;
- lemmatization;
- tokenization;
- word sense disambiguation;
- transform words to vectors using Word2Vec;
- sentiment labeling.

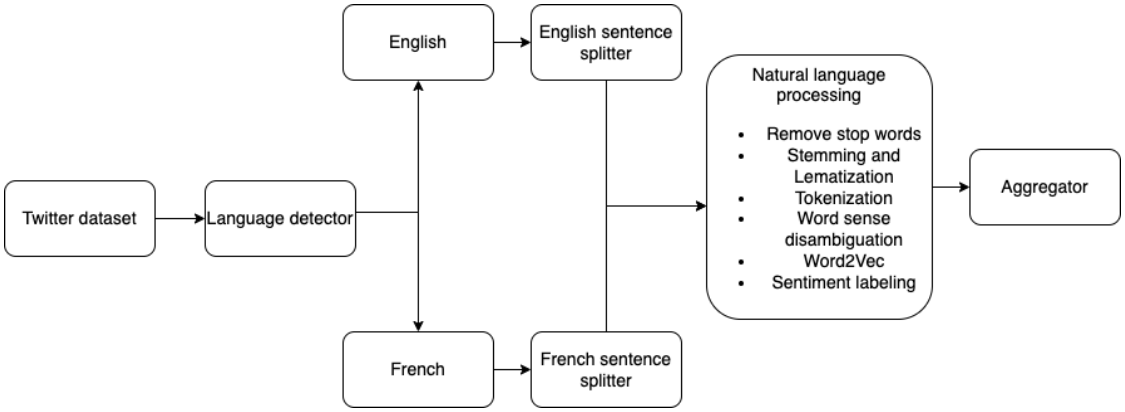


Fig. 2. Detailed diagram of data processing of Twitter CSV dataset

As a result, the algorithm label each tweet from negative to positive represented on a segment from -5 to 5, where -5 is very negative, 0 – neutral, and 5 – very positive. In the final step, tweet scores are aggregated by some period, which could be 1 hour, day, week, month, or custom range inputted by a user.

Data collection. For the research, we collect from Twitter all possible data. The complete list of downloaded data for each tweet is described in Table 1.

Table 1

Downloaded data from each tweet

#	Field	Description
1	Username	Name specified during registration. It is a unique user identifier.
2	Name	No unique name is displayed, which could be edited by a user.
3	Tweet text	The message was written by a user, limited to 280 characters.
4	Hashtags	Proceeds after the "#" symbol. Can be related to an agitation vote for or against any politician.
5	Mentions	Mentioned users. The username proceeds after the "@" symbol.
5	Links	The web address of a website or other tweet.
6	Embedded media	Pictures or videos.
7	Date and time	Date and time when a tweet was published.
8	Replies	Comments to someone's tweet.
9	Favorites	Anyone can highlight tweets that they like.
10	Retweets	Share someone's original tweet.
11	Location	Latitude and Longitude coordinates.
12	Source of the tweet	Android, iPhone or Web.

After collecting tweet data for further research, we need to collect all possible data about the user, who wrote tweets or replies. The complete list of downloaded user data is described in Table 2.

By analyzing user messages and personal information as a result we have a complete portrait of the person and his typical behavior. In most cases, the change in political views is reflected in his social networks acting. For different people it is different. For example, some of them start writing more compliments or hate for political candidates. Some of them just stop writing complement. Such changes are reflected not only in tweets but could also be seen in likes and retweets as well. Some people even change their bio in the profile.

Table 2

Downloaded data for each user

#	Field	Description
1	Username	Name specified during registration. It is a unique user identifier.
2	Name	No unique name is displayed, which could be edited by a user.
3	Location	City, Country.
4	Created date	Date when the Twitter account was created.
5	URL	Link to personal or some other site.
6	Profile Image	User photo or some other image.
7	Language	User-preferred language.
8	Protected	Boolean value (True or False) indicating if a user is protected.
9	Verified	Boolean value (True or False) indicating if a user is verified.
10	Description	Text that the user adds as a profile description.
11	Time zone	Indicates in which time zone the user is.
12	Tweets and replies	All user tweets and replies to other tweets.

Experiment, Results and Discussion

Emmanuel Macron's political rating compared to the sociological poll. The same approach as with the President of Ukraine [10] for political rating detection was tested for the 2021-2022 French presidential campaign. Twitter data was collected weekly from August 2021 until the election itself in April 2022 for two main election candidates – Emmanuel Macron and Marine Le Pen. Overall, the popularity of politicians was growing in social media, candidates had much more new followers compared to those who unfollowed. For example, on August 2021 Emmanuel Macron had 7,219,795 total subscribers but in April 2022 the number became 8,148,825. For each subscriber, the program downloads all tweets from August 2021 and personal data described in Table 1 and Table 2 respectively. By the data, the algorithm found 21,279 bots. Usually, such accounts are created specifically for spreading a lie about politicians, the creation date was in the year 2021 or 2022. Another common part was that such accounts had no real names and surnames in usernames but some text with numbers. By downloading all tweets, we considered which of the accounts have a political position, how often they tweet or like political messages, and whether they changed their opinion over time. After considering and analyzing all possible data political the algorithms forecasted the rating of Emmanuel Macron. Calculated ratings from Twitter data and sociological poll results are presented in Figure 3. Black bars represent the results of the algorithm, and the dashed grey line shows the sociological poll's result (https://www.bbc.com/news/world-europe-59900131). The most important part is that the trend of falling rating right before elections was detected by the proposed algorithm.

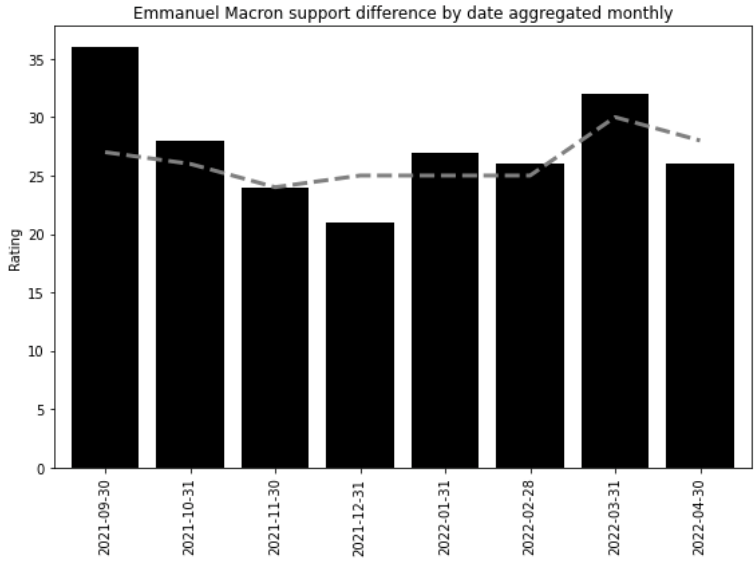


Fig. 3. Emmanuel Macron rating from 30.09.2021 to 30.04.2022

Marine Le Pen's political rating compared to the sociological poll. In the French elections, Marine Le Pen was the main competitor to Emmanuel Macron. Closer to the election date she did a proximity campaign and visited many small towns and villages. Her trips were covered by a lot of local media showing that many people had a chance to see and listen to Marine Le Pen. As with Emmanuel Macron, the popularity in Twitter was growing over time. On August 2021 Marine Le Pen had 2,648,539 total subscribers but in April 2022 the number became 2,818,888. For Marine Le Pen the number of bot accounts was even higher than for Emmanuel Macron – 38 186. Having more bots with a smaller number of subscribers may indicate that more accounts tried to promote Marine Le Pen and

increase the rating of the candidate. Potentially such a strategy may affect rating growth. Calculated Marine Le Pen ratings from Twitter data and sociological poll results are presented in Figure 4.

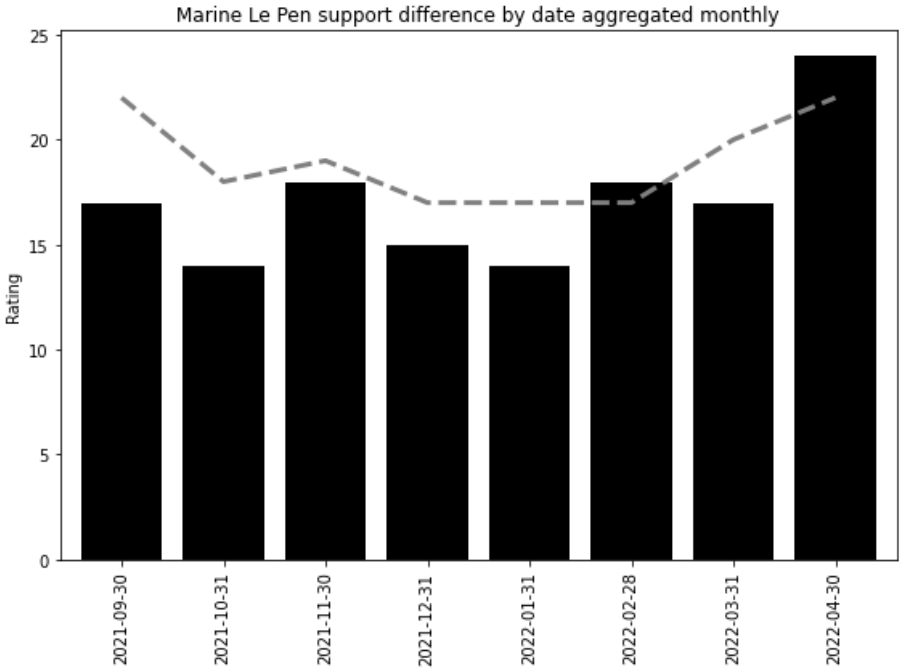


Fig. 4. Marine Le Pen rating from 30.09.2021 to 30.04.2022

Conclusion and Future work

In the conducted research modern types of media are compared to select the best one for election prediction. Selected Twitter social network as the one with the most appropriate data and availability to download big amounts of useful information. The approach based on the usage of Python programming language, Selenium browser emulation and MongoDB database was used to collect, store and clean data about the main French election candidates – Emmanuel Macron and Marine Le Pen. The research was made from August 2021 until the election itself in April 2022. The determined outcome is compared with sociological polls and the results of elections and showed that analysis of social network data could be a good alternative to traditional sociological polls as it shows the same trends month by month and well predicted the win of Emmanuel Macron in elections. Moreover, the proposed approach has its benefits compared to sociological polls such as always being fresh, and close to real-time information, the price of research is much lower and could be reused for the next parliamentary or presidential elections with a small modification.

The research could be extended and adapted for other countries. Currently, the proposed algorithms and mathematical models showed good results in the French and Ukraine elections. It works well with English, French, Ukrainian and Russian languages. This allows us to claim that it will also work fine with other Latin or Cyrillic alphabets but for Asian or Arabic languages more research would be needed. Twitter is a good choice for European and American countries. In the future, other social networks should be considered for the countries in which it is not so popular.

References

1. Y. Roh, G. Heo, S. Euijong Whang. A survey on Data Collection for Machine Learning: A Big Data - AI Integration Perspective. IEEE Transactions on Knowledge and Data Engineering. 2019. Vol. 33. Issue 4. Pp. 1328-1347.

2. H. Liang, J. J. H. Zhu. Big Data, Collection of (Social Media, Harvesting). The International Encyclopedia of Communication Research Methods. 2017. Pp. 397-416.

3. N. Chomsky. Aspects of the Theory of Syntax. MIT Press (1965).

4. C. F. Hockett. Language, mathematics and linguistics. De Gruyter Mouton (1967).

5. S.-H. Liao, P.-H. Chu, P.-Y. Hsiao. Data mining techniques and applications – a decade review from 2000 to 2011. Expert Systems with Applications. 2012). Vol. 39. Issue 12. Pp. 11303–11311.

6. N. R. de Oliveira, P. S. Pisa, M. A. Lopez, D. S. V. Medeiros, D. M. F. Mattos. Identifying Fake News on Social Networks Based on Natural Language Processing: Trends and Challenges. Information. 2021. Vol. 12. No. 38. P. 1-32.

7. C. Sammut, G.I. Webb. TF–IDF. Encyclopedia of Machine Learning. 2011. Pp. 986-987.

8. T. Mikolov, K. Chen, G. Corrado, Jeff Dean, Jeffrey Dean. Efficient Estimation of Word Representations in Vector Space. Computing Research Repository. 2013. arXiv:1301.3781.

9. Q. V. Le, T. Mikolov. Distributed Representations of Sentences and Documents. Computing Research Repository. 2014. arXiv:1405.4053.

10. T. Rudnyk, O. Chertov. Determining Reasons of Political Rating Changes Based on Twitter Data. XXI International Scientific and Practical Conference "Information Technologies and Security". 2021 (in press).

Taras Rudnyk **Тарас Рудник**	PhD Student of the Applied Mathematics Department, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute" https://orcid.org/0000-0001-9492-0374 e-mail: tarasrudnyk@gmail.com	Аспірант кафедри прикладної математики, Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського»
Oleg Chertov **Олег Чертов**	DrSc (Engineering), Professor, Head of the Applied Mathematics Department, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute" https://orcid.org/0000-0003-0087-1028 e-mail: chertov@i.ua	Доктор технічних наук, професор, завідувач кафедри прикладної математики, Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського»

UDC (004.42: 004.93)
https://doi.org/10.31891/csit-2022-4-5

YANA BIELOZOROVA, KATERYNA YATSKO
National Aviation University, Kyiv, Ukraine

FEATURES OF THE IMPLEMENTATION OF THE SPEAKER IDENTIFICATION SOFTWARE SYSTEM

The proposed architecture of the identification software system in the form of class and sequence diagrams. The main criteria for assessing the accuracy of speaker identification were studied and possible sources of loss of speaker identification accuracy were identified, which can be used when building a speaker identification system. A software system based on the proposed architecture and previously developed identification algorithms and methods was created.

The following conclusions can be drawn on the basis of the performed research: approaches to the construction of existing announcer identification systems are considered; the main criteria for assessing the accuracy of announcer identification were investigated and the main sources of loss of accuracy during announcer identification were identified; the structural construction of the announcer identification system is considered, taking into account the identified sources of loss of accuracy during announcer identification; the proposed architecture of the speaker identification system in the UML language in the form of class and sequence diagrams; a software system was built that implements the functions of speech signal identification according to the methods and algorithm proposed in previous works.

The software system uses a ranking method based on three different criteria. These include: calculation of the proximity of two-dimensional probability density function curves for the frequency of the main tone and the location in the spectrum of three frequency ranges that are extracted from the speech recorded in the speech signal; calculation of the proximity of the probability density function curves for each of these features separately; calculation of the degree of closeness of the absolute maxima of the formant spectra extracted from the speech recorded in the speech signal.

Keywords: speaker identification software system, wavelet, diagrams, UML, speech recognition.

ЯНА БЄЛОЗЬОРОВА, КАТЕРИНА ЯЦКО
Національний авіаційний університет

ОСОБЛИВОСТІ РЕАЛІЗАЦІЇ ПРОГРАМНОЇ СИСТЕМИ ІДЕНТИФІКАЦІЇ МОВЦЯ

Обробка мовного сигналу з метою ідентифікації мовця є найбільш актуальною і популярною в задачах, пов'язаних з мовною обробкою. Постійний і високий попит на програмні реалізації систем ідентифікації дикторів існує в різних сферах: від контролю доступу користувачів до голосових послуг виявлення злочинців. Проте, враховуючи відсутність чіткого наукового обґрунтування алгоритмів ідентифікації, значну складність їх реалізації, а також точність ідентифікації особистості, можна відзначити, що ці завдання в цілому ще далекі від свого остаточного вирішення.

Запропонована архітектура програмної системи ідентифікації у вигляді діаграм класів і послідовностей. Досліджено основні критерії оцінки точності ідентифікації мовця та виявлено можливі джерела втрати точності ідентифікації мовця, які можуть бути використані при побудові системи ідентифікації. Створена програмна система на основі запропонованої архітектури та раніше розроблених алгоритмів і методів ідентифікації.

На основі проведених досліджень можна зробити наступні висновки: розглянуто підходи до побудови існуючих систем ідентифікації диктора; досліджено основні критерії оцінки точності ідентифікації диктора та визначено основні джерела втрати точності при ідентифікації диктора; розглянуто структурну побудову системи ідентифікації диктора з урахуванням виявлених джерел втрати точності при ідентифікації диктора; запропонована архітектура системи ідентифікації мовця мовою UML у вигляді діаграм класів і послідовностей; побудовано програмний комплекс, що реалізує функції ідентифікації мовного сигналу за методами та алгоритмом, запропонованими в попередніх роботах.

Програмна система використовує метод ранжування на основі трьох різних критеріїв. До них відносяться: розрахунок близькості двовимірних кривих функції щільності ймовірності для частоти основного тону і розташування в спектрі трьох частотних діапазонів, які виділяються з мови, записаної в мовний сигнал; розрахунок близькості кривих функції щільності ймовірності для кожної з цих ознак окремо; розрахунок ступеня близькості абсолютних максимумів формантних спектрів, виділених з мови, записаної в мовному сигналі.

Ключові слова: програмна система ідентифікації мовця, вейвлет, діаграми, UML, розпізнавання мови.

Introduction

Processing of the speech signal for the purpose of speaker identification is the most relevant and popular in tasks related to language processing. Constant and high demand for software implementations of announcer identification systems exists in various areas from user access control to criminal detection services by voice. However, taking into account the lack of a clear scientific basis for identification algorithms, the significant complexity of their implementation, as well as the accuracy of personal identification, it can be noted that these tasks in general are still far from their final solution.

Related works

The task of automatic speaker verification is considered to be the creation of a mathematical model, a set of algorithms and, as a result of their application, a software or software-hardware implementation that would allow identification of a person with the same accuracy and reliability as is available to a person.

Research efforts in the field of speech technology have led to the appearance of a large number of commercial speech recognition systems. Such companies as Nuance, IBM, ScanSoft offer a large set of software solutions for both server and desktop applications.

To analyze the work of software systems for speaker identification, it is necessary to consider the main approaches to performing evaluations of the work of such systems. The US National Institute of Standards and Technology (NIST) coordinates evaluations of various speech signal analysis systems: automatic speech recognition systems, key word extraction from speech, and speaker recognition. A description of some annual system evaluations can be found in [1]. The Institute develops research methodologies for comparing different systems, which include a clear statement of the task, the definition of evaluation metrics, carefully selected and uniform sets of training and test data for all participants, clear requirements for conducting and providing test results.

For the problems of speech signal identification, there is always a separate decision-making issue [1]. This question should establish the degree of relationship between the declared model and the characteristics of the speech signal being tested. Identification system based on the provided speech signal with language parameters Z person A, must accept one of the following variants of accepted hypotheses:

$H0$: Z defined as A (is taken as a null hypothesis)

$H1$: Z not defined as A

The conclusion regarding the choice of a certain hypothesis is based on the criterion of plausibility based on the assessment of the probability of obtaining differences between the samples:

$$\Delta(Z) = \frac{p(Z|H0)}{p(Z|H1)} \quad (1)$$

where $p(Z|H0)$ and $p(Z|H1)$ - probability density functions (also called likelihood) associated with person A ("own") and "not person" A - («alient»).

In the study of linguistic information identification systems of the well-known US Institute of Standards (NIST), instead of criterion (1), the criterion is recently used [6].

$$\Delta(Z) = \log\left(\frac{p(Z|H0)}{p(Z|H1)}\right) \quad (2)$$

There are many methods for describing a "non-person" model, but we will be most interested in the following two methods. The first method will be based on the selection for each person A of certain standard of templates $\overline{A_1}, \ldots, \overline{A_N}$ [5]. Due to the fact that these templates will be created for each person, it can be concluded that the templates refer to the "non-person" model. The second method is based on the selection as a benchmark of persons falling out of the general distribution, who will correspond to the "not a person" model. This method requires long-term training on test data, but due to adjustment to a large set of input data, it is considered the most effective and is used more often than other methods when building language information identification systems [6].

When developing practical systems for the identification of language information, a certain threshold value θ s most often determined to make a choice about accepting or rejecting a person. The correct choice of θ is always a difficult task that requires a number of experimental studies.

Before using speech signal identification systems, it is necessary to evaluate the accuracy of identification. For such systems, there are three main accuracy factors:

1. crossover error rate (CER);
2. false reject rate (FRR);
3. false acceptance rate (FAR).

A false rejection occurs when the identification system rejects the "correct" option. A false deviation is usually called Type I errors. Such false deviations are quite important in real identification because they lead to a loss of trust in the systems and to the need for additional time for authorization. False acceptance occurs when the "wrong" option is accepted as the "correct" one. False acceptance is also very important, because with this option, an unauthorized person will gain access to resources. It is customary to call a false acceptance Type II errors. Type II errors are thought to be more important than Type I errors because it is better to not admit the real person than to admit the wrong person when making decisions.

The crossover error rate (Fig. 1) is the point of intersection of the error curves of Type I error (FRR) and Type II error (FAR) [7].

The false acceptance rate is also called equal error rate (EER). Considering the fact that this coefficient accumulates the values of false rejection and false acceptance coefficients, it is most often found in the description of speech signal identification systems [8].

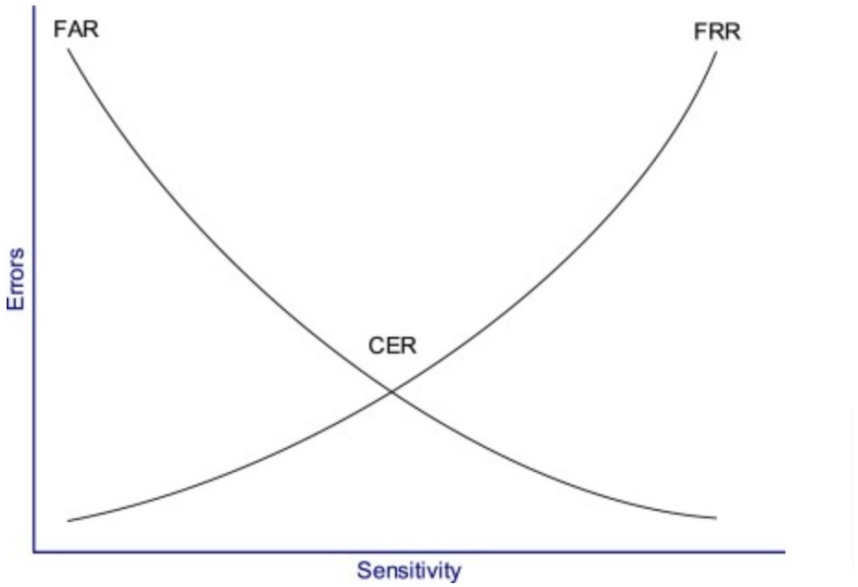


Fig.1. Graph of dependencies of errors of the I and II Types

In the competition among systems of identification of speech signal systems, the learning cost function is also adopted, which determines the weighted sum of the probabilities of false acceptance and false rejection:

$$C_{Det}(\theta) = C_{Miss} \times P_{Target} \times P_{Miss}(\theta) + C_{FalseAlarm} \times \\ \times \left(1 - P_{Target}\right) \times P_{FalseAlarm}(\theta) \quad (3)$$

where the parameters of the cost function are C_{Miss} (missed detection cost factor=1) i $C_{FalseAlarm}$ (false detection cost factor=1), and P_{Target} (a priori probability of the specified target person=0,05), P_{Miss} (probability of missed detection).

Making a decision on the identification of a speech signal is the main indicator and result of research when comparing the speech signals of individuals. Therefore, special attention should be paid to the correct selection of the parameters of the identification system in order to ensure the necessary values of errors of the I and II Types. Most biometric systems have a flexible threshold that controls the balance between these two types of errors. In each program, the optimal threshold is found empirically.

NIST evaluations of various speaker identification systems showed [5]:

– comparison of announcers' voices on the basis of a limited set of data - The point of equality of errors of the I and II types lies within 5-10%. The degree of confidence of the classifier in the obtained result is approximately 95%;

– verification of the announcer based on an extended data set - the point of equality of errors of the I and II types is much lower, in the region of 1.3 - 2%, which roughly corresponds to a relative decrease in the number of errors by 74-80%;

– comparison of announcers' voices based on an extended data set - point of equal probability of errors - 12-15%.

Experiments

In view of the presented identification errors, it can be concluded that the existing announcer identification systems cause fair complaints from users related to the objectivity of examination results. The conducted studies [2] showed that the expressed doubts are fully justified. This conclusion is due mainly to the fact that in most modern means of conducting identification studies of voice signals, the Fourier transformation is used, which is an artificial mathematical method of decomposing a complex signal into periodic components. But the mechanism of perception and transformation of sound vibrations by the human hearing apparatus is arranged differently and such artificial transformations cannot exist in it. It was also established that the main processes of information transmission to the brain, contained in sound signals, are of an impulse nature, and the duration of these impulses ranges from tens to hundreds of milliseconds [4], in connection with the above, it was concluded that the need the use of a multifractal approach to build a speaker identification system [5]. The study of signals in the time domain is necessary because all phonemes have a well-defined fractal character that is preserved and is individual for each phoneme, that is, the form of the phoneme signal in the time domain is the same in all languages and approximately the same when it is pronounced by any individual. It is this uniformity that allows us to recognize the language of any person. The main

difference, which determines the individuality of the speaker, is the individuality of the frequency composition of the signals that make up this sound when it is pronounced by a specific person. This individuality, in our opinion, is determined by the frequency of the main tone and is modulated by the parameters of this frequency. Both the frequency of the main tone and these parameters are determined by the individuality of the components of the vocal tract of any person [2].

During the implementation of the software, two interrelated tasks arose - automatic segmentation of the phonogram and selection, calculation and determination of the degree of proximity of fractal formations contained in the investigated signals of the controversial and exemplary phonograms. Both of these tasks are solved in [5, 6].

Within the framework of the speaker identification task, two interrelated tasks of speaker identification and verification can be distinguished [6]. In the first task, the goal is to identify the audio component as pronounced by one of the announcers from the considered set, in the second - to establish the belonging of the audio component to a specific reference announcer.

Based on these tasks, systems are divided into three parts:

1. determination of individual features of the speech signal;
2. representation of the characteristic standard of the announcer;
3. making a decision about the announcer's personality.

On the basis of the above, it is possible to distinguish the following main stages of the implementation of the announcer recognition system:

Measurement of the fractal dimension of signal components. A stage that is simple to implement, but quite effective in the set of all discriminability measures. Its implementation is possible both with a permanent window and with an adaptive type of window.

Definition of phrase boundaries. To solve this problem, it is most rational to use language segmentation algorithms based on the multifractal approach. Based on this approach, in those elements of the signal, where the change in fractal dimension exceeds some set threshold, it is assumed that a phrase begins.

Selection of the main tone. To solve the task of selecting the main tone, there is a need to develop an interference-resistant method of selecting the main tone for each period. An algorithm based on the use of Morle wavelet approximation of the signal with subsequent statistical analysis of the distribution of wavelet maxima, which is physically explained by the presence of self-similar structures characteristic of signals associated with resonators, can be taken as a basic algorithm for the selection of the main tone.

At the stage of measuring the main tone on the signal sections, it makes sense to compare not the absolute values, but the normalized values - this makes it possible to more accurately distinguish announcers by intonation color.

Selection of characteristic parameters of the main tone. To solve this problem, you can use the finding of only some of the considered parameters during the analysis for each fragment: the average frequency and dispersion of the main tone; distribution of periods of the main tone; amplitude modulation of the main tone; frequency modulation of periods of the main tone.

Comparison of signal parameters with reference parameters. After carrying out the process of comparing speech parameters with the reference ones, you need to select the most "close" speaker from the database. To do this, it is necessary to compare the selected parameters of the main tone from the database based on a probabilistic approach.

In the process of conducting research, the following method of examination was found [4]:

1. To conduct speaker identification research, two or more phonograms are provided for speaker matching, as a rule, in a set of phonograms, at least one phonogram clearly belongs to the voice of a particular speaker.
2. Each phonogram is segmented into fragments based on the fractal dimension [4].
3. For each fragment of each phonogram, the frequency distribution of the main tone over the entire length of the phonogram is calculated based on the frequency distributions of the main tone obtained for the fragments.
4. Phonogram data accurately identifying a person (the given owner of the voice) is stored in the identification database.
5. For each of the phonograms, for which speaker identification must be performed, membership in one distribution for the frequencies of the main tone of each fragment selected from the phonogram is checked, with similar distributions stored in the identification database.
6. Based on the assessment of the degree of closeness between the frequency distribution of the fundamental tone, the announcer is set based on the degree of closeness to the considered distribution.

Let's consider the architecture of the implemented speaker identification system in the UML language in the form of class and sequence diagrams. The class diagram reflects the static structure of the system. It consists of a description of classes and relationships between them. A sequence diagram displays the dynamic relationships in the system, for example, the sequence of calls.

Figure 2 shows a diagram of calls during preliminary preparation for highlighting the characteristics of the announcer's speech. In pre-language recognition mode, the system boots with a prepared configuration file and an input signal. Recognition will be done through the configuration manager.

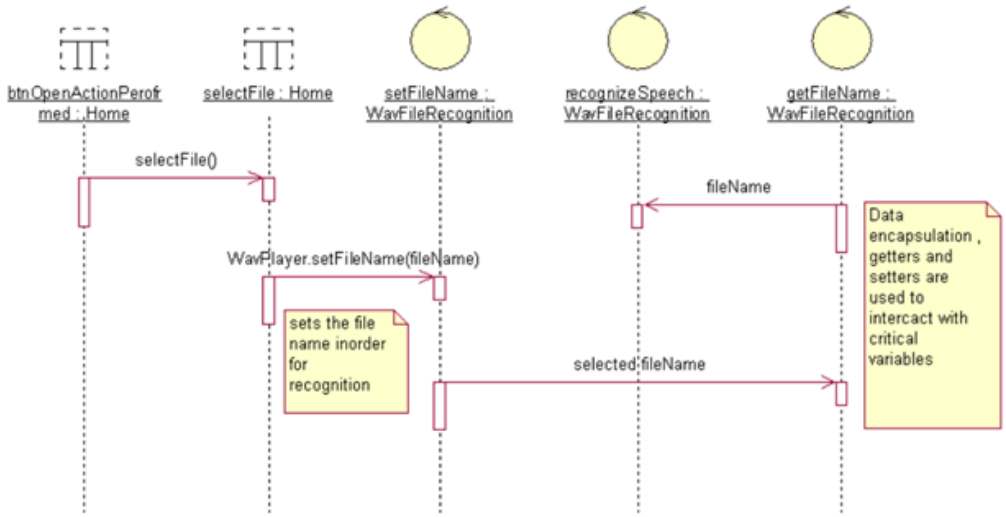


Fig.2. Sequence of preprocessing calls

Figure 3 shows the call diagram during post-processing during speaker recognition. At this stage, the input digital signal will go through the process of dividing it into vocalized and non-vocalized parts, decomposition by Morley wavelet followed by statistical analysis of the distribution of wavelet maxima and determination of the frequency of the main tone for the segments. The AudioFileDataSource and Recognizer classes implement functions to perform these tasks. The result of the sequence of calls are the labels of the class of the announcer and the language to which the classifier classified the input speech signal.

In figure 4 presents a diagram of entity classes, which are object representations of data managed by the identification system.

Home acts as a graphical interface of the software system, which directly interacts with DBSpeaker and WavFileRecognizer. DBSpeaker performs the functions of presentation and description of saved recordings of announcers. WavFileRecognizer is designed to implement the process of reading a sound signal (from a stream or from a file) and identifying the speaker. AudioFileDataSource implements the function of reading the audio signal, and Recognizer implements the speaker identification.

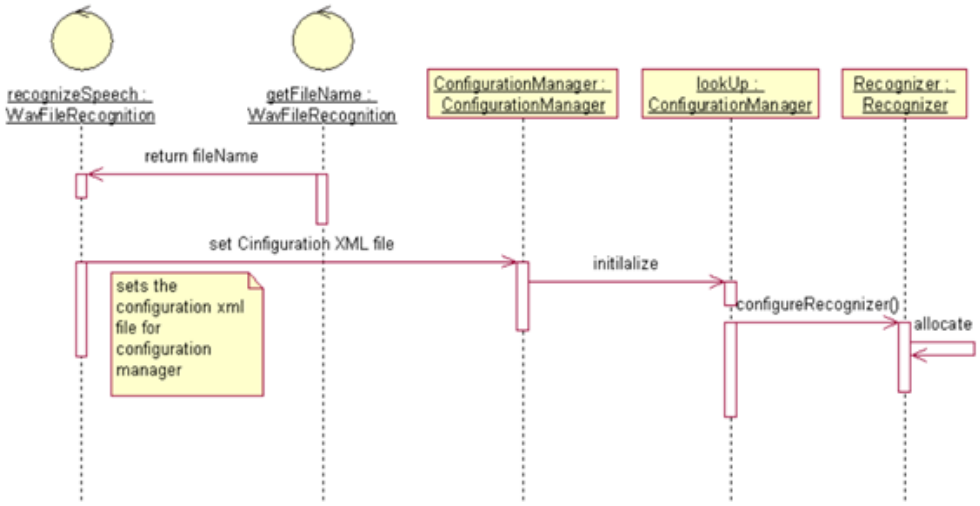


Fig.3. Sequence of calls of the speaker recognition process

The VoiceFeatures abstract class is designed to store and calculate the features of the input speech signal. The class consists of an array of VoiceFeatureValue objects and the ExtractFeatures obtaining method, which performs feature extraction from the input speech signal. The inheritors of the class are the classes performing fragment-by-fragment analysis: average frequency and dispersion of the main tone; distribution of periods of the main tone; amplitude modulation of the main tone; frequency modulation of periods of the main tone.

The PersonClassifier abstract class is designed to implement the classification algorithm. The class consists of the Train and Classify methods, as well as the Parameters object, which contains all the parameters necessary for the work of the classifier. The Train method accepts as input a dictionary, in which the key is a class label, and the

value is an object of type Features, and returns a Parameters object. The PersonClassify method takes a VoiceFeatureValue object and returns the value of the decision feature, as well as the decision class label.

The SpeechUtils class contains helper methods needed for feature computation and classification, such as, for example, computation of vocalized/unvocalized segmentation and denoising.

Thus, a software architecture for speaker identification tasks using a multifractal approach in describing the structure of speech is proposed. The use of a similar architecture and the use of a multifractal approach will generally improve the accuracy of speaker identification.

Based on the proposed architecture, a software system was developed in the Python programming language using the SQLlite database.

The software system for identifying a person's speech signal is searchable, as it is the result of ranking according to the degree of proximity of individual parameters of the speech signal.

The software system for digital recording of informational messages automatically calculates the parameters of language characteristics and further ranks these characteristics in the database of individuals.

The software system uses a ranking method based on three different criteria. These include:

- calculation of the proximity of two-dimensional probability density function curves for the frequency of the main tone and the location in the spectrum of three frequency ranges that are extracted from the speech recorded in the speech signal;
- calculation of the proximity of the probability density function curves for each of these features separately;
- calculation of the degree of closeness of the absolute maxima of the formant spectra extracted from the speech recorded in the speech signal.

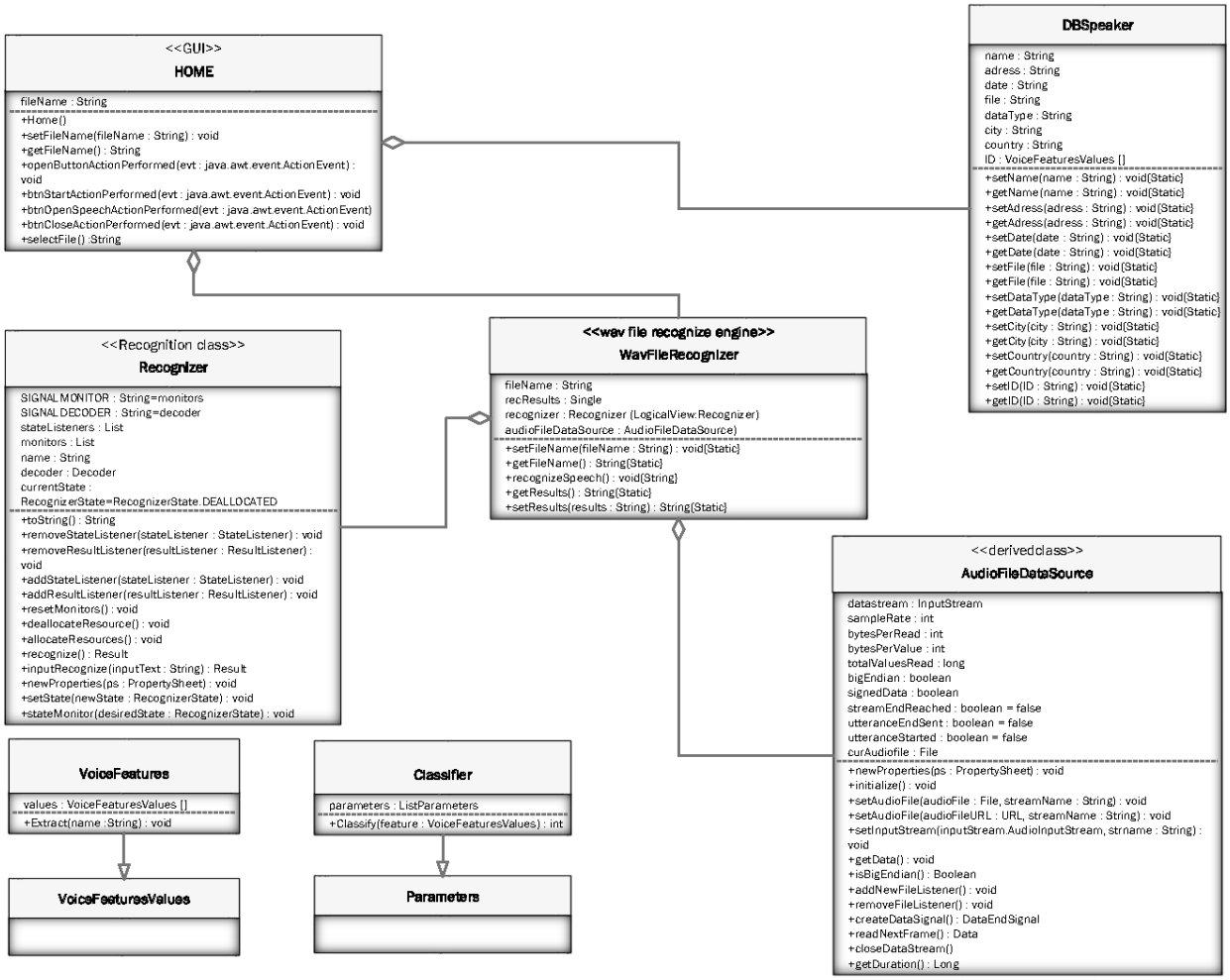


Fig. 4. Class-entity diagram

The result of making a decision of the developed program is a graphical representation of the proximity of the curves of the two-dimensional probability density functions for each of the signs (Fig. 5).

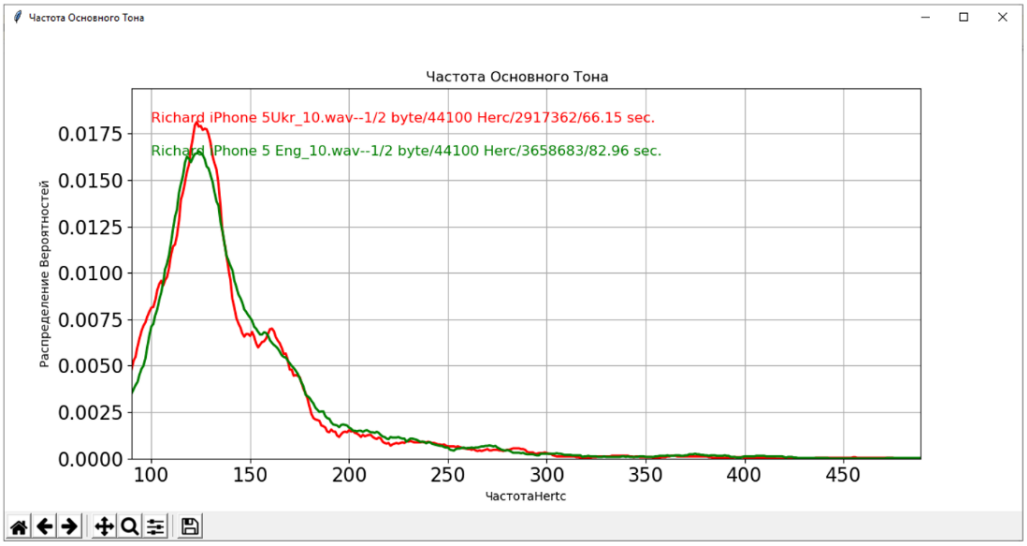


Fig. 5. The resulting graph of the comparison of speech signals of individuals by the frequency of the main tone

Conclusions

The following conclusions can be drawn on the basis of the performed research:

1. approaches to the construction of existing announcer identification systems are considered;
2. the main criteria for assessing the accuracy of announcer identification were investigated and the main sources of loss of accuracy during announcer identification were identified;
3. the structural construction of the announcer identification system is considered, taking into account the identified sources of loss of accuracy during announcer identification;
4. the proposed architecture of the speaker identification system in the UML language in the form of class and sequence diagrams.
5. a software system was built that implements the functions of speech signal identification according to the methods and algorithm proposed in previous works [3-5, 10].

References

1. NIST Speaker Recognition Evaluation. URL:https://www.nist.gov/itl/iad/mig/speaker-recognition
2. Рибальський О.В., В.І. Соловйов, В.В. Журавель, Т.О. Татарнікова. Методика ідентифікаційних і діагностичних досліджень матеріалів та апаратури цифрового й аналогового звукозапису зі застосуванням програмного забезпечення «Фрактал» при проведенні експертиз матеріалів та засобів відео та звукозапису. Науково-методичний посібник : Київ: ДУІКТ, 2013. – 75 с.
3. Solovjov V.I., Byelozorova Ya.A. Multifractal approach in pattern recognition of an announcer's voice. // Teka. Commission of motorization and energetics in agriculture. – 204. – Vol. 15. – № 2. – P. 13-21.
4. Byelozorova Ya.A. The allocation of self-similar structures in voice signals for speaker identification tasks. // ScienceRise. – 2017. – № 5. – P.125-142
5. Бєлозьорова Я.А. Ідентифікація диктора на основі кратномасштабного аналізу// Інженерія програмного забезпечення. – 2017. – №1(29). – С. 15-24.
6. Greenberg S., Singer C., and Mason L. NIST 2020 CTS Speaker Recognition Challenge Evaluation Plan, NIST 2020 CTS Speaker Recognition Challenge, [online], https://www.nist.gov/itl/iad/mig/nist-2020-cts-speaker-recognition-challenge (Accessed June 25, 2021)
7. Hansen J.H.L. A Study on Universal Background Model Training in Speaker Verification / Hansen J.H.L., Hasan T. // Transactions on Audio, Speech, and Language Processing. vol. 19 (7). 2011. pp. 1890–1899.
8. False rejection. URL: https://www.sciencedirect.com/topics/computer-science/false-rejection
9. Татарникова Т.А. Рыбальский О.В., Соловьев В.И., Командина Т.В. Общие подходы к экспертизе оригинальности и подлинности материалов цифровой и аналоговой звукозаписи/ Науковий вісник НАВС. – К.2011, №4, с. 183-191
10. Zybin, S., Bielozorova, Y. Method of Extracting Formant Frequencies Based on a Vocal Signal. In: Hu, Z., Zhang, Q., Petoukhov, S., He, M. (eds) Advances in Artificial Systems for Logistics Engineering. ICAILE 2022. Lecture Notes on Data Engineering and Communications Technologies, vol 135. Springer, Cham. https://doi.org/10.1007/978-3-031-04809-8_40

Yana Bielozorova **Яна Бєлозьорова**	PhD, Associate Professor of Software Engineering Department, National Aviation University, Kyiv, Ukraine https://orcid.org/0000-0002-0688-3436 e-mail: bryukhanova.ya@gmail.com	кандидат технічних наук, доцент кафедри інженерії програмного забезпечення, Національний авіаційний університет, Київ, Україна
Kateryna Yatsko **Катерина Яцко**	Assoc. Prof. of Software Engineering Department, National Aviation University, Kyiv, Ukraine e-mail: kateryna.yatsko@npp.nau.edu.ua https://orcid.org/0000-0002-4389-6033	асистент кафедри інженерії програмного забезпечення, Національний авіаційний університет, Київ, Україна

UDC 004.738
https://doi.org/10.31891/csit-2022-4-6

YAROSLAV KRAINYK
Petro Mohyla Black Sea National University

DEVISING ARCHITECTURE FOR REMOTE EDUCATION ORGANIZATION BASED ON A SINGLE-BOARD COMPUTER

The presented paper sets the goal of providing computer engineering courses based on a single-board computers (SBC) with the organization of the remote education process. The devised approach relies on the connection to the local area network (wired or wireless) of the university, making the board accessible to the internal users, and creation of a network tunnel between the remote user and connected parts. These actions are accompanied by assignments of necessary permissions to the users of the system. First, the requirements of each involved actor were considered and as a result of that further understanding of the network communication process, network management and administration, software requirements, etc. was acquired. According to these results, analysis of the capabilities of SBC in terms of its connectivity was performed and, finally, the architecture was realized on the available hardware with regard to the peculiarities of the specific subject. The article establishes all the necessary details on the implementation of the proposed architecture and discusses results of the testing performed as a part of the specific course. During the testing stage, the deployed system based on FriendlyARM Nano Pi allowed us to organize the course on Real-Time Embedded Computer Systems and successfully finish all related activities under circumstances of remote education. It was demonstrated that the proposed organization is reliable as demonstrates conceivable performance even under condition of multiple simultaneous connections to the target board. From the administrative point of view, the solution is highly configurable which is a huge advantage for the lecturer and for the administrator. The devised architecture has a potential in terms of improvement and scalability because it consumes relatively low network capacity and is easy to deploy under any circumstances.

Keywords: Single-Board Computer, Remote Education, Network, Architecture, Remote Access

ЯРОСЛАВА КРАЙНИК
Чорноморський національний університет імені Петра Могили

РОЗРОБКА АРХІТЕКТУРИ ДЛЯ ОРГАНІЗАЦІЇ ВІДДАЛЕННОГО НАВЧАННЯ НА ОСНОВІ ОДНОНОПЛАТНОГО КОМП'ЮТЕРУ

У представленій роботі запропоновано вирішення питання забезпечення навчальних курсів з напрямку комп'ютерної інженерії на основі одноплатного комп'ютеру (ОК) для можливості віддаленого навчання. Запропонований підхід базується на використанні внутрішньої мережі університету (дротової або бездротової) для того, щоб забезпечити доступ всередині цієї мережі, а також розгортанні мережевого тунелю для віддалених користувачів та пов'язаних компонентів. Ці дії супроводжуються налаштуванням параметрів доступу користувачів до системи. У роботі проаналізовані вимоги усіх сторін, що будуть використовувати та забезпечувати функціонування системи. На основі цього отримані вимоги щодо процесів мережевих комунікацій, мережевого адміністрування, вимоги до програмного забезпечення та інші вимоги. З урахуванням цих результатів проведено аналіз щодо можливостей ОК та реалізовано систему відповідно до отриманих вимог на основі наявного апаратного забезпечення та вимог цільового курсу. Робота представляє деталі відносно реалізації запропонованої архітектури та розглядає результати, які отримані під час розгортання системи. Описана система була розгорнута на базі ОК FriendlyARM Nano Pi для проведення курсу Вбудовані комп'ютерні системи реального часу та успішно забезпечила проходження курсу студентами за умов віддаленого навчання. Продемонстровано, що система є надійною з точки зору продуктивності та за умов одночасного підключення багатьох користувачів. З точки зору адміністрування система надає широкі можливості щодо конфігурування, що є значною перевагою для викладача та адміністратору. Розроблена архітектура має потенціал для подальшого вдосконалення, оскільки вона потребує відносно незначних мережевих ресурсів та може бути легко розгорнута з урахуванням наявних для цього можливостей та вимог і обмежень мережевої інфраструктури.

Ключові слова: одноплатний комп'ютер, віддалена освіта, мережа, архітектура, віддалений доступ.

Introduction

With COVID-19 outbreak pandemic situation, education happened to be in the situation where remote tools and technologies are becoming indispensable for the whole educational process. On-line platforms with multiple features, tools for on-line meetings have gained notable popularity during the last period.

During studying engineering disciplines, the deficiency of such tools is getting to its top extent. Since they suppose direct communication with laboratory equipment, this field of education is one of the most vulnerable to this situation. Experimental stage and work with development boards requires availability of this equipment to students.

The goal of this paper is to establish a solution for remote educational process organization that allows usage of a single-board computer during the remote sessions. It should provide necessary capabilities to ensure that multiple students can work with the equipment simultaneously.

The contribution of this paper is in the following:

– established an architecture for remote educational course organization based on a Single-Board Computer (SBC);

– the architecture incorporates networking and processing features of SBC to organize communication with other parts of the system;

– the proposed solution is scalable and can be applied to the multiple boards deployed in the network of the university.

The rest of the paper is organized as follows. The next section reviews contemporary researches that are concerned with technical aspects of remote education organization. The next third section manifests information technology for single-board computer remote access organization. Results and discussion section exhibits practical experience of the proposed means. The conclusions section summarizes the main points of the proposed information technology.

Related work review

While remote education was under big focus before the pandemic situation [9, 10] and advanced notably in the quality and content, the pandemic situation actually pushed it to one of the top trends in science, social life, and technology.

Remote education is developing in multiple vectors [1, 2, 5, 7]. The means involved in this process are also constantly improving. For instance, the society witnessed a huge success of on-line communication platforms such as Zoom and Google Meet. However, for the case of engineering courses, it is preferable to provide continuous access to hardware without presence of the lecturer. Hence, the solution should not rely only on the communication platforms.

There are many well-known systems developed by the leading universities that establish remote access to the hardware located in the internal network. For instance, in TU Dortmund University, within the development of project ELLI [2], remote and virtual laboratories have been deployed. The main topic of the laboratories is Industry 4.0 and engineering education in general. They maintain access to the industrial equipment and handle control of the available technologies. However, while being an advanced example of remote access to educational environments, it concentrates on the manufacturing process.

The remote laboratory of Maintenance 4.0 [3] exemplifies how remote access may be granted to complex equipment. The solution is based on five-layer architecture with the MIMOSA database. However, this architecture cannot be applied for courses where direct communication with hardware is preferable. Administration and programming tasks should be executed on the hardware platform.

The ArPi project [4] has deployed a low-budget educational system based on Raspberry Pi computer and Arduino boards with Ethernet shields integrated into the local network. The system has sophisticated API for communication and a huge set of connected devices. The control system is implemented using hyper-text protocol and is accessible as a web-project. However, ArPi Lab project mainly focuses on the work with Arduino controllers. Raspberry Pi plays the role of a server and cannot be programmed during the course. Since single-board computers are more powerful devices, it is preferable to have additional access to such devices.

The analysis of the actual curriculum of the leading universities in the field of computer engineering demonstrates that they presume work with single-board computers as one of the options for hardware platforms. SBC provide necessary balance of convenient software infrastructure and low-level access to the hardware interfaces [8]. Moreover, as they use Linux-based operating systems predominantly as a main software layer, the transition from desktop software development to development of software for peripheral devices should be smooth and fluent.

One of the latest trends in education is a burst of popularity of virtual reality (VR) [6] and augmented reality (AR) tools. They provide improved digital experience for situations when visual experience is important. However, in case of engineering courses that suppose work with hardware basis and software development, use of AR or VR may be impractical. Thus, students need the technology that allows them to connect to the hardware remotely and execute programming actions.

Therefore, the problem of maintaining computer engineering courses based on SBC with extended access to hardware is still relevant. Remote access to the hardware is the key component for such courses organization. The rest of the paper is dedicated to the solution of this problem using available infrastructure of the university.

Information technology and architectural solution for remote access organization

It is a common case that universities have complex internal network infrastructure that is protected and not accessible from the outside except public resources due to security reasons. When the student has access to the network, all educational resources deployed inside should be completely available. However, as those resources are not accessible from the outside, it poses a huge problem to keep the educational process stable for all participants of engineering courses.

Recently single-board computers (SBCs) like Raspberry Pi, Orange Pi, Nano Pi, etc. have achieved extreme popularity in the engineering educational environment due to their powerful features alongside with affordable price. In fact, many of them are powerful enough to perform complex computations. They facilitate support of novel software libraries and can communicate with external devices. Therefore, they are highly exploited in many university courses.

First, let us explain how SBC may be used in the educational process. The use-case diagram is shown in Fig. 1.

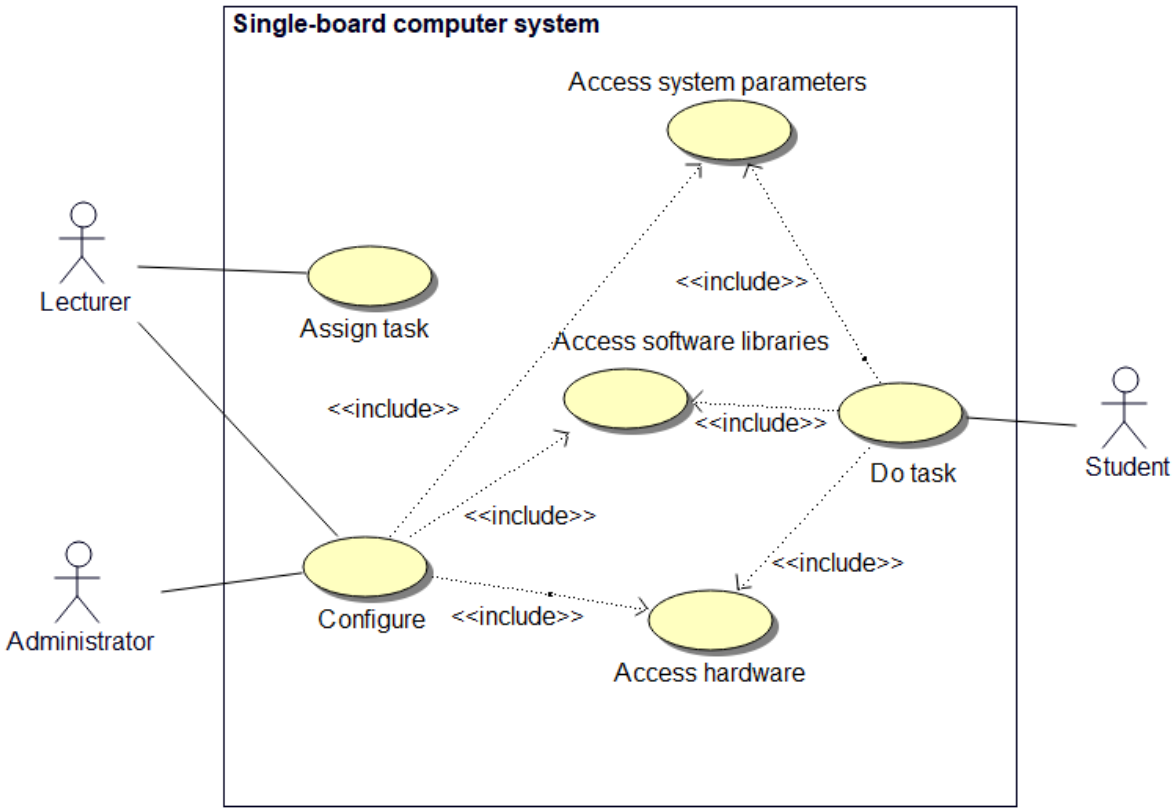


Fig. 1. Use-case diagram of the devised system

Three roles are present in the system. The lecturer assigns a task and has access to the configuration parameters. The lecturer also can configure the system to prepare it for students. This task may be delegated to the administrator. Notice that execution of "Access..." use-cases from the student's side is not the same as from administrator's side because of different levels of permissions. For example, while administrator and lecturer has complete access to hardware, the student is limited to its limited set and cannot execute critical functions. The lecturer can install any software while the student can only use software and libraries that are installed. The student's permissions are restricted to the level required to execute assigned tasks.

Local network integration. To provide remote access to the SBC hardware, first, it needs to be integrated into the local network of the university. Generally, university LAN provides both wired and wireless connections. The first one creates a basis for network distribution on the complete campus. The wireless utilizes resources of the wired network and provides connectivity for user devices (smartphones, laptops, etc.). They rely on Ethernet and Wi-Fi technologies respectively. Hence, there are two options on how to connect SBC to the internal infrastructure. The decision about the medium to connect with should be done according to the requirements to the reliability of connection and throughput. Available interfaces are also to be taken into consideration because many SBCs do not have an Ethernet port.

Hereinafter, three different cases are considered:

– wired connection;
– wireless connection to the secured network;
– wireless connection to the ad-hoc network.

The last two options may demonstrate different situations as even one university student can have both variants depending on the network architecture.

Wired connection. In this case, a conventional configuration file /etc/network/interfaces needs editing. Policy for retrieving IP-address (DHCP or static address), gateway, and network mask are the minimum parameters set to prepare.

To configure connection to the Wi-Fi, it is possible to use wpa-supplicant tool. Its main goal is to simplify the process of configuration of wireless interfaces and automate it. The most convenient way to pass settings to the utility is to prepare a configuration file.

Configuration file for wpa-supplicant tool has a format shown in Fig. 2. First, let us examine configuration for an ad-hoc network.

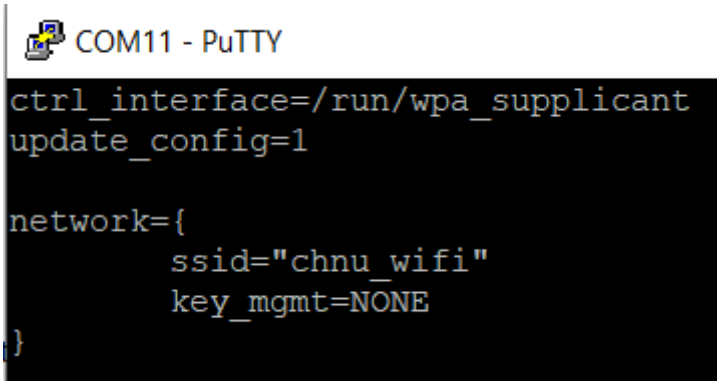


Fig. 2. Content of WiFi configuration file

Ad-hoc network supposes that the user does not need to enter credentials to access its resources. They are quite widespread in various social institutions. It is also quite common for universities to provide such services. As can be observed in Fig. 3, the value of the key management setting is assigned to NONE so no credentials will be used during connection.

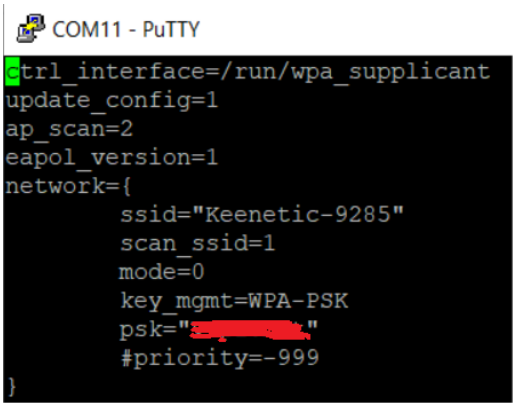


Fig. 3. Configuration for a secured network

To apply the setting during the startup stage, the wireless interface of the SBC has to be prepared by the following lines in /etc/network/interfaces (Fig. 4).

```
allow-hotplug wlan0
iface wlan0 inet dhcp
        wpa-conf /etc/wpa_supplicant/wpa_supplicant.conf
```

Fig. 4. Excerpt from the network interfaces configuration file

The last line executes configuration stored in the .conf-file.

If a DHCP system is used for IP-address assignment, the address of the board can be reserved exclusively to the SBC to avoid continuous changes. Thus, the address will be known in advance.

Access to the board. The access to the board is established using a common protocol for remote access, Secure Shell (SSH). Therefore, the port 22 must be open for external connections. As long as additional protocols are required, they may be installed by the administrator. To be able to interact with the SBC, the user needs credentials. Due to the security reasons credentials for each permitted user is generated before the start of the course when it is confirmed that he/she attends the course. In this way, no external students will consume resources of the SBC.

In general, the students receive a numerical identifier and associated password to access all university electronic resources. It is convenient to use the numerical identifier to grant permissions to the SBC. To automate this task the script that generates credentials and creates system users should be available to the administrator and the lecturer.

Besides being able to login and execute general commands, the student of the engineering course is supposed to work with hardware components. For instance, access to the I2C or UART by default has only the user with administrator privileges. Without additional configurations, the ordinary user cannot read or write data to/from the device. The permissions can be granted in the following way. First, the list of resources should be identified. Second, each item of the list should be checked for vulnerabilities from unexperienced user's actions. Third, the logging system should be turned on to collect information about all system actions so the possible source of the problem may be

detected by logs analysis. As soon as those conditions are met, the administrator can traverse all required resources and provide access to all users or the specific group of users of the system using chmod command. Once again, to automate this task all commands may be placed into a script that runs during the startup.

On the other side, not all settings and permissions can be assigned in advance because the system is dynamic and new instances of devices/modules may appear during the board work. For instance, controlling pins through the virtual file system sysfs leads to creation of directories with restricted permissions. In this case, the lecturer provides permissions by request.

Software libraries. Apparently, students should not have permissions to install new software to the SBC. On the other hand, it is a vital point of multiple learning steps to have an opportunity to write the software part and execute it. Hence, at the preliminary stage, all necessary libraries and programs (compilers, script languages, etc.) should be installed and tested. Typically, the standard setup process configures correct access to the utilities to all groups of users. In opposite, this can be configured by the administrator during the preparation stage.

Remote access. Previously, the remote access to the SBC from outside of the university has been outlined as the main problem. To ensure remote access for students, features of the tunneling protocol are exploited. Point-to-Point Tunneling Protocol (PPTP) is the protocol that allows remote access organization to the dedicated resources in the internal infrastructure. It requires creation of a Virtual Privacy Network (VPN) with PPTP as a communication protocol. Actually, the student connects to a dedicated server that also runs VPN to accept clients' connections. As soon as connection is provided, the client can access internal network resources as if they are within the same network. Hence, the scheme of remote access organization is shown in the Fig. 5.

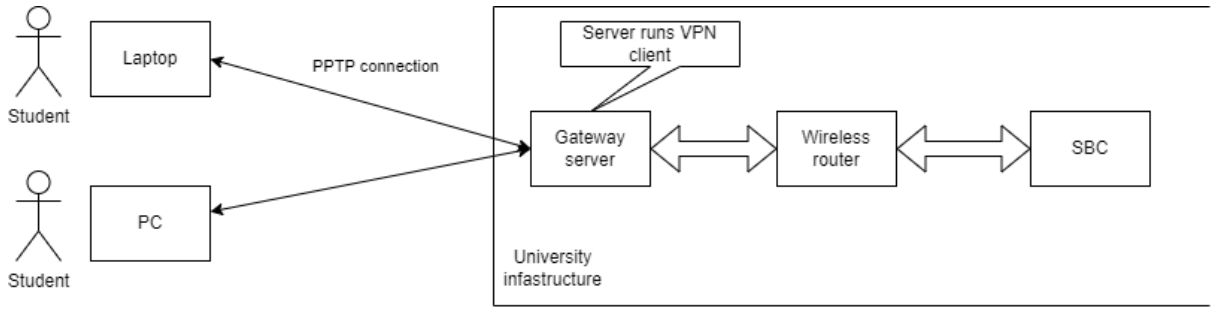


Fig. 5. General network communication scheme for the proposed architecture

Notice, that in order to secure the server, the administrator of the internal network should generate credentials and associate them with the user that has no privileges to harm the whole system in any capacity. Otherwise, the system becomes vulnerable to external attacks. Additionally, to decrease the load of the server, the administrator may limit the number of simultaneous connections.

As soon as VPN is configured, the client can connect to the SBC via SSH-protocol. The lecturer or administrator can execute necessary activities by being on premise (e.g. using UART) as well as having access to the SBC from the remote location.

Results and discussions

During the experimental stage, Nano Pi Neo Air board was used. FriendlyARM Nano Pi Neo Air is a contemporary SBC with rich set of characteristics. It supports WiFi-connection but has no port for wired connection. It has enough memory (512 Mb RAM) and computational resources (quad-core with frequency up to 1.2 GHz) to ensure simultaneous work of a students' group. Because of his small form-factor (40x40 mm) it consumes almost no space and its installation may be done practically anywhere providing a reliable level of WiFi-signal.

All configurations have been done using serial connection to the board. Since there are two wireless networks in Petro Mohyla Black Sea National University (Mykolaiv, Ukraine) that match both described cases, connection was prepared for both networks.

The access to the SBC was provided to the students of Real-Time Embedded Computer Systems (master level). During the course they managed to do laboratory tasks with internal infrastructure of the SBC, connected components, software programming and debugging. The brief list of tasks that were assigned to the students included the following items:

- Introduction to the file system organisation, device identification in the file system;
- Identification of the available commands and their location;
- Work with preinstalled utilities to communicate with a six-axis sensor via I2C interface;
- Deployment of the dummy driver and is demodulation;
- Compilation and execution of a software program that controls connected LEDs or processes measurements from the above-mentioned sensor;
- Usage of the debugging capabilities of the system.

The following peak values of the Nano Pi parameters were observed under condition of approximately 20 simultaneous connections to the SBC:

- RAM usage - 50%;
- CPU usage - 83%.

As demonstrated the values, the peak observed value was relatively high, however, most of the time CPU resources are occupied not more than 10% of the maximum value. The measurements were received from the calls of cron service that initiated logging of the specific parameters into a log file. Additionally, those numbers prove that the solution provides necessary level of scalability despite constraints of the SBC device. While adding one more SBC to the network by the same scenario does not imposes significant difficulties, the benefits of having multiple instances can ensure better load balancing for the resources of SBC deployed.

Conclusions

In this paper, the architecture for remote educational infrastructure organization for engineering educational courses based on SBC usage has been established. The architecture allows complete remote access to the hardware located within the university network. The deployed instance of Nano Pi Neo Air was used during the testing stage. The results demonstrate that students can have access to the computer and have sufficient experience to do tasks remotely.

References

1. Kans M., Campos J., Hakansson L. A remote laboratory for Maintenance 4.0 training and education // IFAC-PapersOnLine, 2020. – Vol. 53, Iss. 3. – P. 101-106. DOI: https://doi.org/10.1016/j.ifacol.2020.11.016
2. Grodotzki J., Ortelt T.R., Tekkaya E. Remote and Virtual Labs for Engineering Education 4.0: Achievements of the ELLI project at the TU Dortmund University // Procedia Manufacturing , 2018. – Vol. 26. – P. 1349-1360. DOI: https://doi.org/10.1016/j.promfg.2018.07.126
3. Gullu A., Aki O., Kuscu H. Remote Access for Education and Control of Mechatronics Systems // Procedia - Social and Behavioral Sciences, 2015. – Vol. 176. – P. 1050-1055. DOI: https://doi.org/10.1016/j.sbspro.2015.01.577
4. Kaluz M., Cirka L., Valo R., Fikar M. ArPi Lab: A Low-cost Remote Laboratory for Control Education // IFAC Proceedings Volumes, 2014. – Vol. 47, Iss. 3. – P. 9057-9062. DOI: https://doi.org/10.3182/20140824-6-ZA-1003.00963
5. Ugur M., Savas K., Erdal H. An internet-based real-time remote automatic control laboratory for control education // Procedia - Social and Behavioral Sciences, 2010. – Vol. 2, Iss. 2. – P. 5271-5275. DOI: https://doi.org/10.1016/j.ifacol.2015.11.212
6. Esquembre F. Facilitating the Creation of Virtual and Remote Laboratories for Science and Engineering Education // IFAC-PapersOnLine, 2015. – Vol. 48, Iss. 29. – P. 49-58. DOI: https://doi.org/10.1016/j.ifacol.2015.11.212
7. Xu H., Sun Y. Remote classroom system for Chinese linguistics teaching based on FPGA and embedded system // Microprocessors and Microsystems, 2021. – Vol. 81. DOI: https://doi.org/10.1016/j.micpro.2020.103785
8. Post L.S., Guo P., Saab N., Admiraal W. Effects of remote labs on cognitive, behavioral, and affective learning outcomes in higher education // Computers & Education, 2019. – Vol. 140. DOI: https://doi.org/10.1016/j.compedu.2019.103596
9. Bawa P. Learning in the age of SARS-COV-2: A quantitative study of learners' performance in the age of emergency remote teaching // Computers and Education Open, 2020. – Vol. 1. DOI: https://doi.org/10.1016/j.caeo.2020.100016
10. Garg S., Diwakar A., Upadhuay S.K., Kumar G., Singh G. EFFECT OF COVID-19 ON SCHOOL EDUCATION SYSTEM: CHALLENGES AND OPPORTUNITIES TO ADOPT ONLINE TEACHING AND LEARNING // Humanities & Social Sciences Reviews, 2020. – Vol. 8(6). – P. 10-17. DOI: 10.18510/hssr.2020.862

Yaroslav Krainyk **Ярослав Крайник**	PhD, Associate Professor, Head of Computer Engineering Department, Petro Mohyla Black Sea National University, Mykolaiv, Ukraine https://orcid.org/0000-0002-7924-3878 e-mail: yaroslav.krainyk@chmnu.edu.ua	кандидат технічних наук, доцент, завідувач кафедри комп'ютерної інженерії, Чорноморський національний університет імені Петра Могили, Миколаїв, Україна.

UDK 004.4
https://doi.org/10.31891/csit-2022-4-7

OKSANA KYRYCHENKO
Yuriy Fedkovich Chernivtsi National University (Ukraine)

INFORMATION TECHNOLOGY FOR STATISTICAL CLUSTER ANALYSIS OF INFORMATION IN COMPLEX NETWORKS

Information technology has been developed, which is used to collect, process and save large volumes of data from the web space. With the help of technology, the statistical characteristics of various segments of the web space and their cluster structure are studied. Two methods are used to find the optimal number of clusters and cluster centers: the well-known k-core decomposition algorithm and a new method developed by the authors. The new algorithm is based on the distribution of eigenvalues of the stochastic matrix, which describes the process of Markov transitions in the system. The clustering process is carried out using the Power iteration clustering algorithm.

With the help of written software (crawler), information is collected on a given segment of the web space. For the studied area, there are statistical characteristics, namely: node degree, clustering coefficient, node probability distributions by input and output connections. Oriented and unoriented graphs of web pages of the studied zones are constructed. By combining the calculated dependencies for the input and output subnets, we can obtain the statistical characteristics of the undirected graphs of the web pages of the web space zones that we are investigating.

For cluster analysis, the optimal number of clusters and cluster centers can be found in 2 ways: by the well-known k-core decomposition algorithm and by using a new method developed by the author. The new algorithm is based on the distribution of eigenvalues of the stochastic matrix, which describes the process of Markov transitions in the system. Using the Rower iteration clustering algorithm, the cluster structure of various segments of the web space is studied.

The advantage of the developed information technology is that with its help one can work with large sets of data collected on the Internet, study their structure and statistical characteristics, and perform the clustering process. To implement the clustering process and find the optimal number of clusters and centroids a new algorithm is suggested. The results of the algorithm indicate high accuracy in determining the optimal number of clusters.

Keywords: optimal number of clusters; cluster centers; k-core decomposition algorithm; eigenvalues; stochastic matrix; clustering process; statistical characteristics, process of Markov.

ОКСАНА КИРИЧЕНКО
Чернівецький національний університет ім. Ю. Федьковича

ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ ПРОВЕДЕННЯ СТАТИСТИЧНО-КЛАСТЕРНОГО АНАЛІЗУ ІНФОРМАЦІЇ У СКЛАДНИХ МЕРЕЖАХ

Велика кількість інформації в Інтернеті та й загалом сам інформаційний простір являють собою складну мережу з усіма характерними для таких структур статистичними характеристиками, особливостями та зв'язками. Вивчення статистичних особливостей і кластерної структури таких мереж, а також найбільших доменів і зон цікавить сьогодні багатьох дослідників і вчених.

Розроблена інформаційна технологія, за допомогою якої проводиться збір, обробка та збереження даних великих об'ємів з веб-простору. За допомогою інформаційної технології досліджуються статистичні характеристики різних сегментів веб-простору та досліджується їх кластерна структура.

За допомогою написаного програмного забезпечення (кроулера) проводиться збір інформації по заданому сегменту веб-простору. Для досліджуваної зони знаходяться статистичні характеристики, а саме: ступінь вузла, коефіцієнт кластерності, розподіли ймовірностей вузлів за вхідними та вихідними зв'язками. Будуються орієнтовані та неорієнтовані графи веб-сторінок досліджених зон. Об'єднуючи розраховані залежності для вхідних та вихідних підмереж, можна отримати статистичні характеристики неорієнтованих графів веб-сторінок зон веб-простору, які досліджуємо.

Для проведення кластерного аналізу знаходиться оптимальне число кластерів та центри кластерів 2 способами: відомим алгоритмом k-core decomposition та за допомогою нового методу, розробленого автором. Новий алгоритм базується на розподілі власних значень стохастичної матриці, що описує процес Маркова переходів у системі. За допомогою алгоритму Power iteration clustering проводиться дослідження кластерної структури різних сегментів веб-простору.

Перевагою розробленої інформаційної технології є те, що з її допомогою можна працювати з великими масивами даних, зібраних в Інтернеті, вивчати їх структуру та статистичні характеристики, здійснювати процес кластеризації. Для реалізації процесу кластеризації та знаходження оптимальної кількості кластерів і центроїдів запропоновано новий алгоритм. Результати роботи алгоритму свідчать про високу точність визначення оптимальної кількості кластерів.

Ключові слова: оптимальне число кластерів, центри кластерів; алгоритм k-core decomposition; власні значення; стохастична матриця; процес кластеризації; статистичні характеристики, марковський процес.

Introduction

A large amount of information on the Internet, and in general the information space itself, represent a complex network with all the statistical characteristics, features and connections typical for such structures. The study of statistical features and cluster structure of such networks, as well as the largest domains and zones, is of interest to many researchers and scientists today.

In the theory of complex networks, three main directions of research are considered:

• studying the statistical characteristics that specify the behavior of networks;

• creating a network model;

• predicting the network behavior when its structure changes.

Active development of this field of research led to the study of network characteristics, taking into account not only its topology, but also statistical characteristics that characterize the behavior of the network when the structural properties change.

For this, researchers study the statistical characteristics of various networks: energy networks, transport networks, air transport networks, computer networks, co-authorship networks, social networks, the Internet and many others [1-5].

Many works are devoted to studying the structure of the WWW space and statistical characteristics of the web space [1-5]. The structure of such networks is presented by graphs. The nodes of the graph are the web pages, and the edges are the links between them. Both directed and undirected graphs are studied [2-4]. It was found that the World Wide Web obeys the statistical laws of complex networks, and it was established that the distribution of the nodes of the graph, which reflects the World Wide Web, obeys a power law with the indicator close to (-2,2) for input connections and (-2, 7) – for output ones [2-4]. This indicates the scalelessness of such a network, i.e. the high level of development of the network as a whole [1-4].

Description of information technology

The goal of our research consisted in the development of information technology for collecting, processing, saving and conducting statistical cluster analysis of information in complex networks. It is possible to study a complex network in several stages. The structure of the developed information technology is shown in Fig. 1.

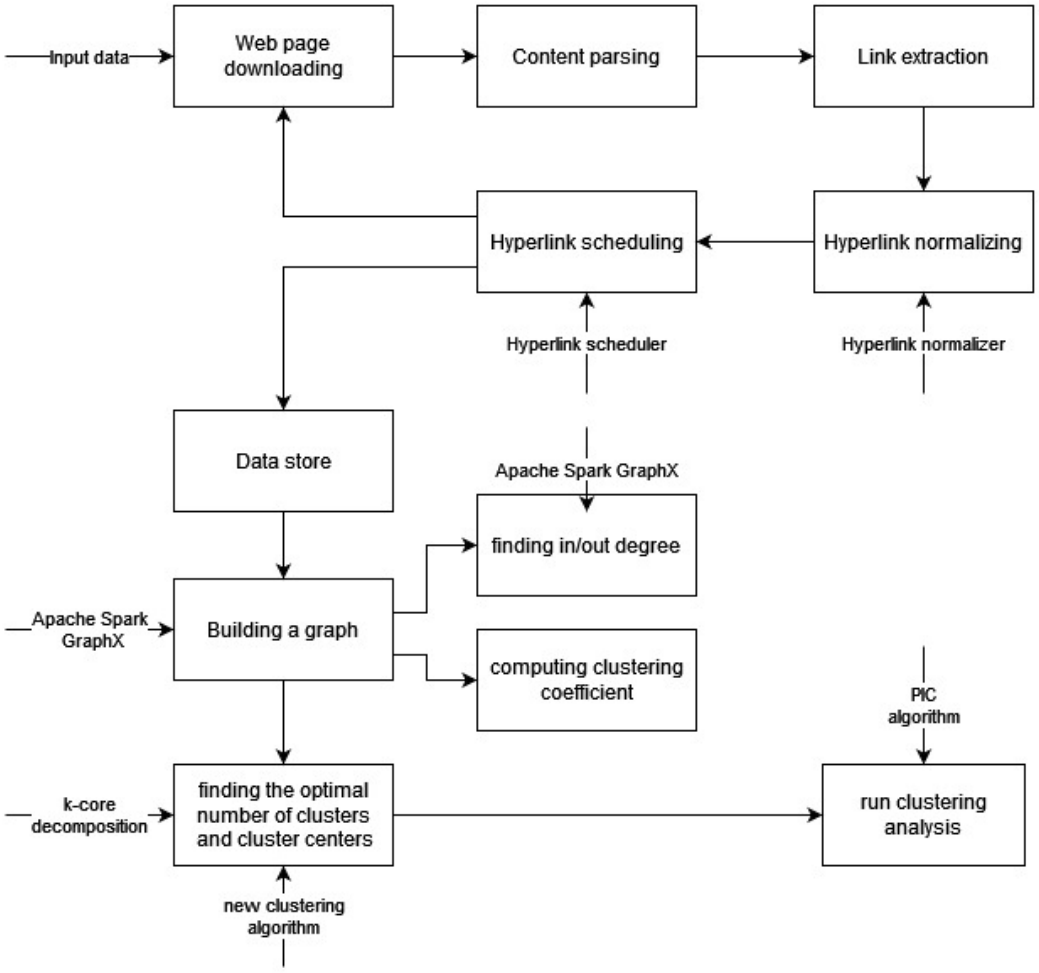


Fig.1. General scheme of information technology for conducting statistical cluster analysis of information in complex networks

We have developed the software (crawler) that is completely controlled by the settings before starting work. The user specifies a list of web page addresses – entry points, and if necessary, indexing depth, etc. This enables to fully control the process of searching and indexing pages as well as the calculating the main statistical parameters of the network under study [6]. The program was configured for a given number of "jumps" from the entry point on its links. The results of the research are recorded to the database, and the crawler moves on to the next entry point. If because of "wandering" the crawler comes across a page that is already in the database, no repeated study is required,

only a new connection is added. As a result, we obtain a graph of web pages, the statistical characteristics of which will be studied. For the area under study, the degree of each node is found, the clustering coefficient is determined, and probability distributions of nodes are constructed based on input and output connections. Combining the calculated dependencies for the input and output subnetworks, the statistical characteristics of the undirected graphs of the web pages of the investigated web space zones can be obtained.

Thus, a graph is built and the clustering process should be carried out. The process of clustering enables to consider large data sets and drastically reduce their dimensionality, make them compact and explore their structure. The task of clustering consists in dividing a set of objects into groups of similar objects, called clusters. The result of the clustering process is a set of clusters that contain similar elements of the input set of elements.

Concerning the clustering task, there arises a problem of choosing a clustering algorithm: it should be chosen in such a way that the division into clusters be the most correct.

One of the most common problems with clustering algorithms is that for most of them the number of clusters as an input parameter is required; however, the number of clusters is usually not known in advance. That is why the use of certain empirical rules to choose the optimal number of clusters is needed.

A large class of clustering algorithms is based on representing the sampling as a graph. The vertices of the graph correspond to the objects of the sampling while the edges correspond to the pairwise distances between the objects. An adjacency matrix is constructed, which is then examined. Spectral clustering is often used to divide a set of large dimensions into clusters. The data dimension is reduced using the spectral clustering algorithms. The next step consists in applying some clustering method (e.g., k-means). It should be noted that the main drawback of algorithms based on the k-means method is the requirement to determine the initial number of clusters and their centers.

Information about these parameters is usually not available at the initial stage of the information space research.

Another disadvantage of the k-means method is that it does not provide a solution to the problem of determining the optimal number of clusters in the data sampling.

After the stage of graph building, we determine the optimal number of clusters and find the cluster centers. Such process can be performed in two ways:

• by the k-Core decomposition method [7];

• by a new method developed by the authors, which is based on the asymptotic distribution of the eigenvalues of a stochastic random matrix without conditions of element independence, the spectrum of which can be decomposed into a regular part and outliers [8].

The next step is to perform the clustering process, i.e. the division into clusters, which is carried out using the Power Iteration Clustering (PIC) algorithm [9].

Unlike the spectral clustering algorithms, the PIC algorithm calculates only one eigenvector (which is actually a linear combination of several eigenvectors). In this way, a high speed of calculations is achieved if compared to the traditional spectral clustering algorithm [9]. The PIC algorithm pseudocode is given below:

Input: Normalized similarity matrix W, number of clusters k
Output: Clusters $C_1, C_2, ..., C_k$

1. Pick an initial vector $\mathbf{v}^0$.
2. $\mathbf{v}^{t+1} \leftarrow \frac{W\mathbf{v}^t}{\|W\mathbf{v}^t\|_1}$ and $\boldsymbol{\delta}^{t+1} \leftarrow |\mathbf{v}^{t+1} - \mathbf{v}^t|$.
3. Increment t and repeat above step until $|\boldsymbol{\delta}^t - \boldsymbol{\delta}^{t-1}| \simeq \mathbf{0}$.
4. Use k-means on $\mathbf{v}^t$ and return clusters $C_1, C_2, ..., C_k$.

Performing such tasks results in division of the collected data set into clusters and the possibility to compare the performance of different clustering algorithms.

k-core decomposition method

To determine the optimal number of clusters, we used the k-Core decomposition method.

A k-Core graph is a maximally connected subgraph in which every vertex is connected to at least k vertices in the subgraph. The k-core distribution is often used in largescale network analysis [7]. Its main aim is to find a strong subgroup, the members of which play the role of communicators on the graph. Each node in the subgraph must have at least k degree.

k-Core decomposition has the following properties:

$$\forall u \in V\colon k\text{-}core\,(u) = k \leftrightarrow$$

$$\begin{cases} \text{There is such maximal subgraph } V_k \text{ that } \forall v \in V_k\colon deg(v) \geq k, \\ \text{and} \\ \text{There is no such subgraph } V_{k+1} \text{ that } \forall v \in V_{k+1}\colon deg(v) \geq k+1. \end{cases}$$

This algorithm consists in finding the subgraph with the strongest connections *k*. This means that each member of this subgraph has at least *k* neighbors. In addition, there is no larger subgraph where each member has more than *k* neighbors. Therefore, if we find a vertex that has the highest degree in this subgraph, it will be a good candidate for its cluster center [7].

New clustering method

Along with the classical clustering method, for example k-means, in our approach we will use a new spectral clustering method, in which the selection of clusters is based on the transition matrix (eigenvalues) *P*. In [8, 10 – 12], the asymptotic distribution of the eigenvalues of the stochastic matrix *A* with random elements a_{ij} □ *dist*, where *dist* is the basic distribution of elements. It was determined that the normalized matrix with the elements (1)

$$p_{ij} = \frac{a_{ij}}{\sum_{j=1}^{N} a_{ij}}, \tag{1}$$

corresponds to the asymptotic distribution of the real parts of the eigenvalues of the matrix *P* with a distribution density

$$f_{\lambda}(x;N) = c\sqrt{(a_{+}-x)(x-a_{-})}, \quad x \in (a_{-}, a_{+}),$$

where a_{-}, a_{+} – distribution parameters determined by the following relations (2)

$$a_{\pm} = \frac{\sigma^2}{\sqrt{N}}, \tag{2}$$

c – normalizing constant, N – the number of elements (elements in the network) in the matrix *P*. At the same time, it was established that the statement about the distribution of eigenvalues has a circular distribution.

Moreover, the number of clusters is suggested to be chosen according to the following rule

$$k_{opt} = \#\left\{\lambda_i : |\lambda_i(P) - 1| \le \frac{1}{\alpha\sqrt{N}}\right\}, \tag{3}$$

where the constant *α* depends on the intercluster connection in the matrix *A*.

Process of technology testing

This information technology was used to conduct a statistical cluster study of the following areas of the web space: the Polish segment of the web space (edu.pl), the Israeli segment (ac.il) and the Ukrainian (net.ua and edu.ua). For each segment of the web space, the probability distribution of nodes according to incoming connections (in degree) and the probability distribution of nodes according to degrees by outgoing connections (out degree) are constructed. The average values of a node degree for undirected graphs were constructed and determined [13]. The graphs of the investigated web space zones were built. The network clustering coefficients were calculated (Table 1). The data indicate a large number of the nearest-neighbors cross-references [13].

Table 1.

Clustering coefficients for subnetworks

Name of zone	edu.ua	net.ua	ac.il	edu.pl
Clustering coefficient	0.11	0.13	0.104	0.088

We can see that for all networks the clustering coefficients are within the value of 0.1, which also indicates similar statistical characteristics of all studied segments. It can be concluded from the research results that according to their statistical characteristics the studied segments of the web space (net.ua; edu.ua – Ukrainian, ac.il – Israeli and edu.pl - Polish) belong to scale-free networks. This fully corresponds to the modern trends in the development of the Internet.

The study of the cluster structure of the obtained data sets began with determining the optimal number of clusters using two described methods: the k-Core decomposition and a new spectral method developed by the authors

[8]. Based on the Monte Carlo methods it was found that the new spectral algorithm is more accurate in estimating the number of clusters. The networks under study were divided into clusters using the PIC algorithm.

Conclusion

The advantage of the developed information technology is that with its help one can work with large sets of data collected on the Internet, study their structure and statistical characteristics, and perform the clustering process. To implement the clustering process and find the optimal number of clusters and centroids a new algorithm is suggested. The results of the algorithm indicate high accuracy in determining the optimal number of clusters.

References

1. Broder A. Graph structure in the web / A. Broder, R. Kumar, F. Maghoul et al. // Proceedings of the 9th World Wide Web Conference, Computer networks, 2000. - 33 (1). – P. 309-320.
2. Newman, M.E.J. The Structure and Function of Complex Networks. *SIAM Review*. – 2003. – Vol. 45. – N. 2. – P. 167–256.
3. M. E. J. Newman. The structure of scientific collaboration networks. Proc. Natl. Acad. Sci. USA 98, 404–409 (2001).
4. S. H. Strogatz. Exploring complex networks. *Nature* 410, 268–276 (2001)
5. Newman, M. E. J. Models of the small world: a review. J. Stat. Phys. 101, 819–841 (2000).
6. Kyrychenko O.L., Kanovsky I., Ostapov S.E. Software for the statistical characteristics of the global WWW network research. *Information Processing Systems*, 2013, Vol. 2(3), pp. 99-104.
7. Sheng-Tzong Cheng, Yin-Chun Chen, and Meng-Shuan Tsai, 'Using k-Core Decomposition to Find Cluster Centers for k-Means Algorithm in GraphX on Spark', CLOUD COMPUTING 2017: The Eighth International Conference on Cloud Computing, GRIDs, and Virtualization.
8. Kyrychenko O.L., Malyk I.V., Ostapov S.E. Cluster structure analysis of Internet networks based on random matrixes. *International Scientific Technical Journal "Problems of Control and Informatics",* 2022, 1, pp. 37-46.
9. Frank Lin and William W. Cohen, 'Power iteration clustering', in ICML(to appear), (2010).
10. Robert C. Qiu, Paul Antonik (2017). Smart Grid using Big Data Analytics. A Random Matrix Theory Approach. Wiley Online Library, 2017.
11. T. Tao and V. Vu. Random matrices: Universality of the local eigenvalue statistics. *Acta Math*. 206 (2011), no. 1, 127–204.
12. T. Tao and V. Vu. Random matrices: universality of local eigenvalue statistics up to the edge. *Comm. Math. Phys.* 298 (2010), no. 2, 549–572.
13. Oksana Kyrychenko, Sergey Ostapov, Igor Kanovsky. Comparison of Statistical Characteristics of Certain Internet Subdomains. Monograph. Scientific Publishing of the Academy of Business in Dabrowa Gornicza: Wydawnictwo Naukowe, 2014. – 138 p. (ISBN: 978-83-62897-91-9)

Oksana Kyrychenko **Оксана Кириченко**	Assistant Professor, Department of Mathematical Problems of Control and Cybernetics, Yuriy Fedkovych Chernivtsi National University, Chernivtsi, Ukraine, e-mail: o.kyrychenko@chnu.edu.ua https://orcid.org/0000-0003-0282-9958	асистент кафедри математичних проблем управління і кібернетики, Чернівецький національний університет імені Юрія Федьковича, Чернівці, Україна.

UDC 004.27; 621.382
https://doi.org/10.31891/csit-2022-4-8

LARYSA MARTYNOVYCH, YURII GUNCHENKO,
YURII SHUGAILO, YURII BERCOV
Odesa I.I. Mechnykov National University, Odesa, Ukraine

DESIGN AND SYNTHESIS OF TERNARY LOGIC ELEMENTS

The aim of this paper is creating some ternary elements. The threshold element of ternary logic on bipolar transistors and elements of ternary systems based on it are considered. The main disadvantages of this approach are identified. The multi-threshold element of multivalued logic and its specific four-threshold implementation are considered. The scientific nowelty is the using of a multithreshold element of multivalued logic as a basis for constructing elements of ternary systems. It is shown that the advantages of a multi-threshold element of multi-valued logic are a larger number of thresholds of input signals, a larger number of levels that it can distinguish and a larger number of output signals. The implementation of some ternary elements, such as half-adder, disjunction and strong conjunction, is given. The practical significance of obtained results is that the multithreshold element of multivalued logic allows us to use it for synthesis basic elements of ternary logic and use one approach for all of them. Additional practical advantage is that in practice it allows you to build more diverse logical and arithmetic devices with a simplified implementation. The implementation of ternary devices based on threshold logic is a way to create ternary devices that can compete with binary devices in terms of equipment, capacity, operational capabilities, and variety. This, in turn, leads to greater speed and simplification of the structure of devices, as well as increasing the speed of data processing. These advantages are significant in such areas of computer use as intelligent data processing systems, expert systems, decision theory, i.e., where data analysis is performed. Prospects for further research are to use the method of construction and synthesis of nodes of ternary computer systems, based on MTEML. Their optimization, and development of principles of mathematical modeling and software of such systems and their elements is expedient.

Keywords: ternary logic, threshold element of ternary logic, multi-threshold element of multi-valued logic, disjunction, strong conjunction

ЛАРИСА МАРТИНОВИЧ, ЮРІЙ ГУНЧЕНКО,
ЮРІ ШУГАЙЛО, ЮРІЙ БЕРКОВ
Одеський національний університет імені І.І.Мечникова,

ПРОЕКТУВАННЯ ТА СИНТЕЗ ТРІЙКОВИХ ЛОГІЧНИХ ЕЛЕМЕНТІВ

Метою роботи є створення деяких трійкових елементів. В роботі розглянуто пороговий елемент трійкової логіки на біполярних транзисторах та елементи трійкових систем на його основі. Визначено основні недоліки такого підходу. Розглянуто багатопороговий елемент багатозначної логіки та його конкретну чотирипорогову реалізацію. Наукова новизна полягає у використанні багатопорогового елемента багатозначної логіки як основи для побудови елементів трійкових систем. Показано, що перевагами багатопорогового елемента багатозначної логіки є більша кількість порогів вхідних сигналів, більша кількість рівнів, які він може розрізнити, і більша кількість вихідних сигналів. Наведено реалізацію деяких трійкових елементів, таких як напівсуматор, диз'юнкція та сильна кон'юнкція. Практичне значення отриманих результатів полягає в тому, що багатопороговий елемент багатозначної логіки дозволяє використовувати його для синтезу базових елементів трійкової логіки та використовувати один підхід для всіх них. Додаткова практична перевага полягає в тому, що на практиці це дозволяє будувати більш різноманітні логічні та арифметичні пристрої зі спрощеною реалізацією.

Реалізація трійкових пристроїв на основі порогової логіки – це спосіб створення трійкових пристроїв, які можуть конкурувати з двійковими пристроями за кількістю обладнання, потужністю, експлуатаційними можливостями та різноманітністю. Це, в свою чергу, призводить до більшої швидкості та спрощення структури пристроїв, а також збільшення швидкості обробки даних. Ці переваги є значущими в таких сферах використання комп'ютерів, як інтелектуальні системи обробки даних, експертні системи, теорія прийняття рішень, тобто де проводиться аналіз даних. Перспективами подальших досліджень є використання методу побудови та синтезу вузлів потрійних комп'ютерних систем на основі MTEML. Доцільною є їх оптимізація, розробка принципів математичного моделювання та програмного забезпечення таких систем та їх елементів.

Ключові слова: трійкова логіка, пороговий елемент трійкової логіки, багатопороговий елемент багатозначної логіки, диз'юнкція, сильна кон'юнкція.

Introduction

Ternary logic for computers, introduces its units of information: trit and trite (like binary bit and byte). Trit (approximately 1,585 bits) is a ternary digit in the ternary number system. Trite is the minimum ternary word addressed in the memory of a ternary computer. Often, one trite is equal to six trits.

In digital electronics, the "bit" is implemented by the minimum logical element of a binary computer - a binary trigger. Trit is implemented by a ternary trigger, which can simultaneously operate with three values at once, instead of two as in the binary trigger. One trite can encode 729 values at 6 trits, against 256 one byte at 8 bits. It takes values from the range of -364 to 364, in contrast to the byte range 0 - 255. This allows you to process much more information in one CPU clock [1].

The practical feasibility of ternary technology is not yet clear. The ternary technique is equivalent to the binary technique in the sense that everything feasible in one of them, with one or another approximation can be done in another. It is also clear that ternary elements must be more complex and expensive than binary, and ternary logic is clearly more complex than binary. But, on the other hand, and more diverse, ternary memory is more powerful, and the operational capabilities of ternary valves are richer.

Data processing in the conditions of ternary technology should be carried out at the same physical speed of the elements faster, and the structure of the ternary device, as a rule, is simpler than the structure of a functionally equivalent binary device. In other words, the ternary technique is characterized in comparison with the binary complication of the elements, which makes it possible to simplify the structures created from them, reduce the number of communication lines, and increase the speed of data processing.

Thus, it is obvious that ternary technology and systems based on it are more efficient and convenient than binary in intelligent systems, in expert systems, in decision making theory.

Related works

To date, the element base of ternary systems, as well as the ternary systems themselves, is at the stage of its development. Various solutions and methods for constructing ternary elements are proposed [2]. There are different approaches and principles of their construction [3]. Consider some of them.

Threshold element on bipolar transistors

The implementation of a ternary logic element on bipolar complementary unsaturated transistors - a threshold element of ternary logic (TETL) is known [4]. The element is implemented based on a binary ECL-element, the circuit of which is supplemented by its replica on complementary transistors. TETL consists of a block of emitter repeaters (BER) and connected to its outputs m blocks of current switches (CS.1… CS.m). BER is implemented on two repeaters, respectively on n-p-n and p-n-p transistors. The first repeater is connected between the common bus and the power bus "-E", the second - between the power bus "+ E" and the common bus. Each CS unit contains 2 current switches. Fixed currents If are formed by two current sources connected in accordance with the power buses "+ E" and "-E".

The disadvantage of TETL is that a small number of thresholds is determined, because of which the levels "++" = "0+" and "- -" = "0-" do not differ.

However, the circuit and structural solutions used in them and tested in practice can find application in modern digital technology.

Devices based on TETL

Some devices of ternary logic are realized on the basis of TETL, such as the former of three-digit constants, ternary repeater, non-cyclic inverter, the scheme "OR".

The device for forming constants (FC) has 2 outputs: +1 and -1 [5,7]. At any input ternary value, the outputs are constantly exposed to +1 and -1 (discrete currents of a fixed value + If and -If). The purpose of FC is to set a constant triple value at the right points of the scheme: +1 and -1. In this way, the desired point of the device is tied to a given logical value.

The two-input circuit "OR" has two outputs - direct and inverse.

Let`s consider in detail a ternary half-adder, which was implemented on the threshold elements. The corresponding truth table is presented in table 1.

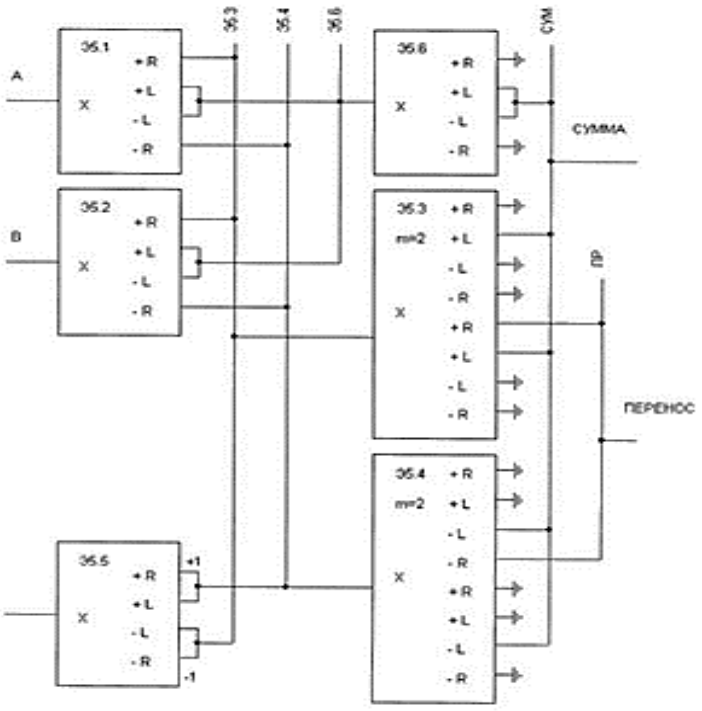


Fig. 1: Triple half-adder on the basis of TETL

This circuit contains five identical elements with one block of current switches (m = 1): E5.1, E5.2, E5.3, E5.4, E5.5 and two identical elements with two blocks of current switches (m = 2): E5.3, E5.4 as it`s shown on Fig. 1. Elements E5.1 and E5.2 divide the ternary variables A and B into their two-digit components. Element E5.5 is used as a constant shaper.

For the final formation of the SUM signal, a double inverted TRANSFER signal is fed to the SUM bus, which is formed by combining the signals of the same name + L and –L from the outputs of elements E5.3 and E5.4. For doubling (formation of signals + 2If and -2If) in elements E5.3 and E5.4 use two current switches and two sets of outputs (m = 2).

Table 1

Truth table for ternary half-adder

№	a	b	S	C
1	-	-	+	-
2	-	0	-	0
3	-	+	0	0
4	0	-	-	0
5	0	0	0	0
6	0	+	+	0
7	+	-	+	0
8	+	0	0	0
9	+	+	-	+

a, b - input signals, S - the sum of the input signals, C - overflow, positive for two positive input signals, and negative - for two negative input signals.

Also, a node of ternary circuitry (Fig. 2), consisting of 3 PETL was built on the basis of TETL [5].

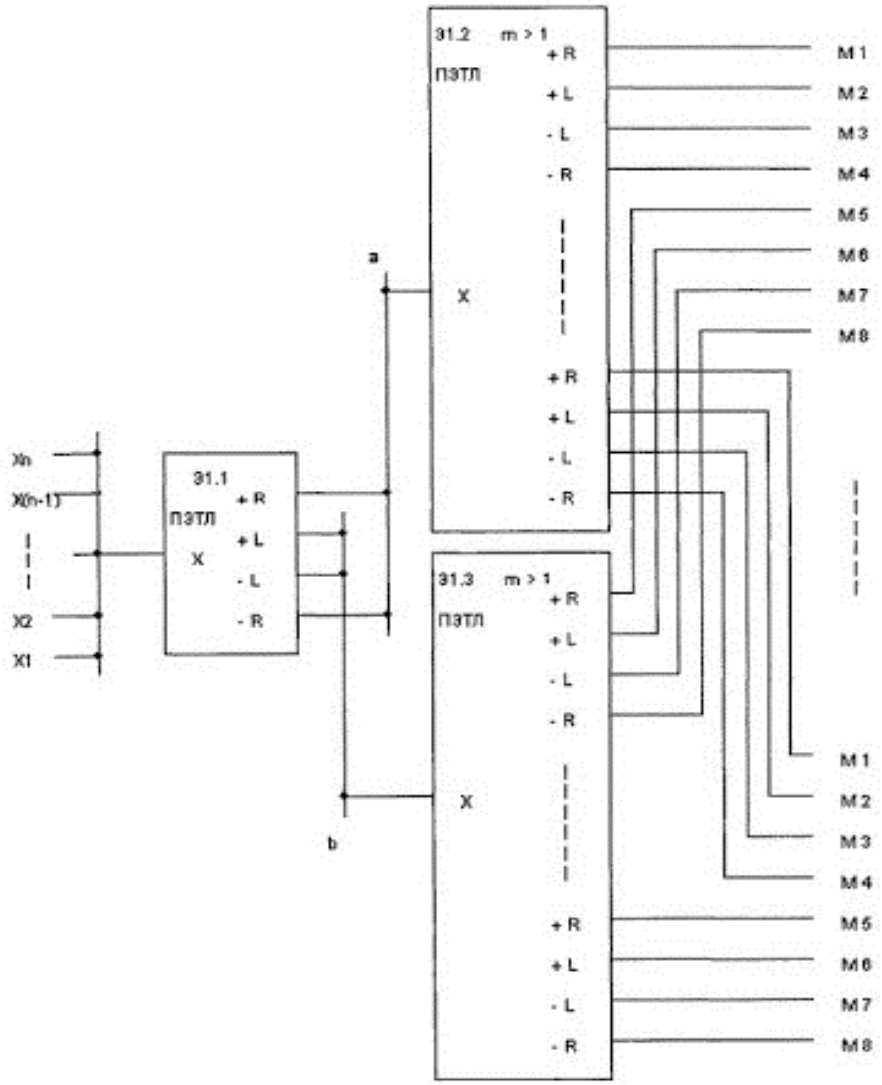


Fig. 2: Node of ternary circuitry

Another device built on the basis of TETL is a triple reversing shift register.

The nodes of this element consist of a TETL and for them everything related to the TETL is valid: the inputs algebraically sum up the triple values; nodes can have more than one group of outputs [6].

Triple Reverse Offset Register operates with data presented in a ternary code with numbers +1, 0 and -1, and implements the following functions:

- storage of k-bit ternary values;
- parallel writing and reading;
- consecutive writing and reading;
- shift right and left by a specified number of digits.

Considering the devices built on the basis of TETL, we can conclude that all these elements do not allow the full implementation of ternary logic, do not have a common approach to its implementation, and complicate the implementation of ternary devices and their structure. Therefore, the issue of developing a standard approach and methods for the synthesis of ternary elements is quite relevant.

Proposed technique

Multithreshold element of multivalued logic

To solve this problem, let's consider a multi-threshold element of multivalued logic [8] as a basic element for constructing elements of multivalued systems, including ternary. Its block diagram is shown in Fig. 3.

Consider in detail its structure and principle of operation. Block 1 is a block of threshold formation (BTF), 2.1… 2.n - emitter repeaters, 3.1 … 3.m - current switches. The input of the threshold formation unit receives k discrete current signals Ij from the previous elements. They can take one of the typical values (for example, for the binary logic of such values will be two: Ij = +1, Ij = 0; for a ternary symmetric system of such values will be three: Ij = +1, Ij = 0, Ij = –1).

$$k = k_{+1} + k_{-1} + k_0,$$

where k_{+1} – the number of signals whose current values +1,
k_{-1} – the number of signals whose current values -1,
k_0 – the number of signals whose current values 0.

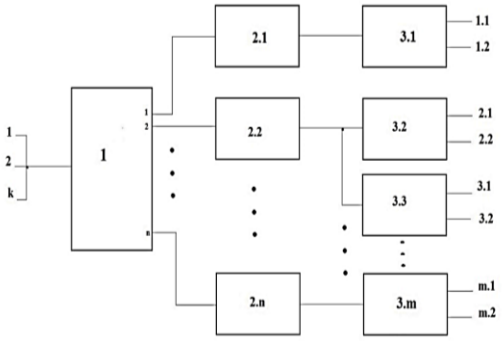


Fig. 3: Block diagram of MTEML

BTF forms n thresholds in one turn. Its outputs are fed to the inputs of emitter repeaters (ER) 2.1… 2.n. Active ER generate signals on the connected current switches (CS) 3.1… 3.m.

Features of the structure of MTEML: the element does not operate with potential, but with current values of signals, so the outputs of MTEML can be combined in any number, but the signal can be applied only to the input of one element; ability to form any number of thresholds that MTEML can distinguish.

The number of BTF thresholds depends on the number of levels of the input variable that the MTEML is able to distinguish and, accordingly, the bit size of the variable or the complexity of the operations that can be performed. In order to form the logic of the operation of this element, you need to combine the outputs of the CS in the required combination.

Four-threshold implementation of MTEML

Consider an example of the implementation of MTEML for a ternary symmetric system - its four-threshold version [9], the block diagram of which is shown in Fig. 4.

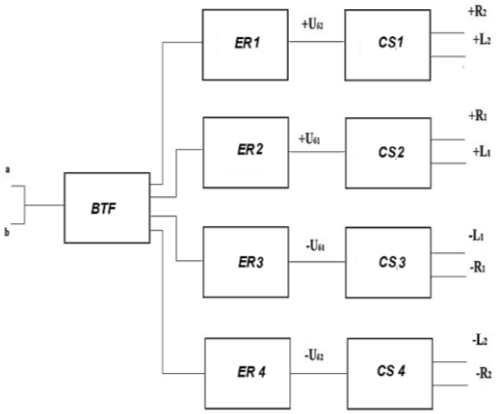


Fig. 4: Block diagram of the four-threshold MTEML

The values of voltages generated at the outputs of the ER and respectively fed to the inputs of current switches are given in table. 2. In this table "1" is an active signal (affects the CS), "0" is an inactive signal (does not affect the CS) at the output of the corresponding ER.

Table 2

Voltages generated at the output of emitter repeater blocks

The sum of the input currents (terlev)	ER1 (+ $U_{б2}$)	ER2 (- $U_{б2}$)	ER3 (+ $U_{б1}$)	ER4 (- $U_{б1}$)
– –	1	0	1	0
–	0	0	1	0
0	0	0	0	0
+	0	0	0	1
++	0	1	0	1

If the signal at the input of the CS is active, the output L of the corresponding CS current is generated, otherwise - the current is generated at the output R. CS together have 8 outputs, a combination of which can form the necessary logical or arithmetic functions [10].

The values of the output signals of the CS depending on the input currents are described by the terlev function, given in table 3.

Table 3

The values of the outputs (output currents) of the current switches CS1 – CS4

The sum of the input currents	Output signals of current switches							
	CS 1		CS2		CS3		CS4	
terlev	+R2	+L2	–R2	–L2	+R1	+L1	–R1	–L1
– –	0	+	–	0	0	+	–	0
–	+	0	–	0	0	+	–	0
0	+	0	–	0	+	0	–	0
+	+	0	–	0	+	0	0	–
+ +	+	0	0	–	+	0	0	–

The implementation of MTEML for ternary symmetric logic, in comparison with TETL, has several advantages: 4 thresholds of input signals (in TETL - 2), distinguishes 5 levels (in TETL - 3), has 8 output signals (in TETL - 4). All this together allows you to build more diverse logical and arithmetic devices with a simplified implementation.

Experiments

Elements of ternary systems based on a multi-threshold element of multivalued logic

Triple half-adder

Based on MTEML the ternary half-adder was constructed [11] which structural scheme is shown in fig. 5.

The input of the BFT 1 receives two ternary variables a and b. BFT forms four symmetrical thresholds. The outputs of the BFT are fed to the emitter repeaters ER1 - ER4, at the output of which voltages are formed, depending on the input function terlev, which are given in table 2.

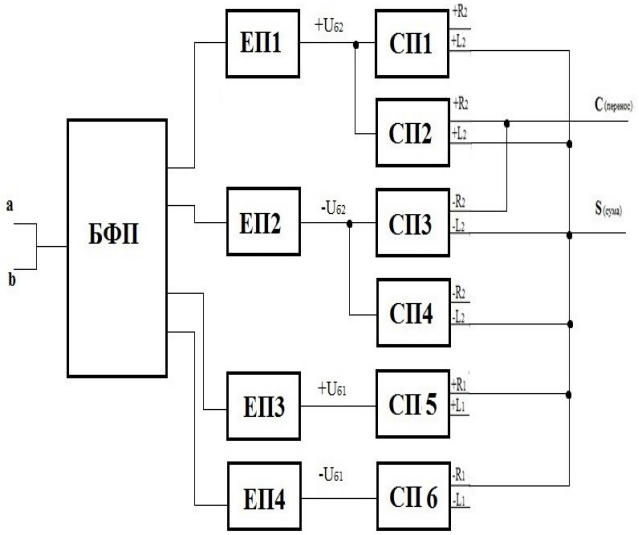


Fig. 5: Block diagram of a ternary half-adder based on MTEML

The table 4 shows the values of the outputs of the CS depending on the function terlev. This function illustrates the transformations performed by a multi-threshold element of multivalued logic, based on which a ternary half-adder is constructed [12].

Using table 4, it is possible to construct operations of addition on the module 3 - S and transfer - C (table 5).

Table 4

The values of the outputs (output currents) of the current switches CS1 - CS6

The sum of the input currents	Output signals of current switches							
	CS 1, CS2		CS3, CS4		CS5		CS6	
terlev	$+R_2$	$+L_2$	$-R_2$	$-L_2$	$+R_1$	$+L_1$	$-R_1$	$-L_1$
– –	0	+	–	0	0	+	–	0
–	+	0	–	0	0	+	–	0
0	+	0	–	0	+	0	–	0
+	+	0	–	0	+	0	0	–
+ +	+	0	0	–	+	0	0	–

From the table 4 it follows that :

$$S = 2(+L_2(ab), -L_2(ab)) + (+R_1(ab), -R_1(ab))$$
$$C = (+R_2(ab), -R_2(ab))$$

The constructed ternary half-adder contains fewer elements and has a simplified circuit design solution, due to the fact that it is built on the basis of MTEML.

Table 5

Truth table of ternary half-adder

a	b	terlev	S	$2(+L_2(ab), -L_2(ab))$	$+R_1(ab), -R_1(ab)$	C	$+R_2(ab), -R_2(ab)$
–	–	– –	+	++	–	–	–
–	0	–	–	0	–	0	0
–	+	0	0	0	0	0	0
0	–	–	–	0	–	0	0
0	0	0	0	0	0	0	0
0	+	+	+	0	+	0	0
+	–	0	0	0	0	0	0
+	0	+	+	0	+	0	0
+	+	++	–	– –	+	+	+

In order to demonstrate this, it is necessary to bring the TETL-based half-adder to a form similar to the constructed ternary half-adder (Fig. 6).

The naked eye can see from the diagram that the structure of the half-adder based on the TETL is much more complex. The number of elements in it is much larger. As well as many unused outputs.

The structure of the ternary half-adder on the basis of TETL consists of six threshold elements of ternary logic, which in turn contain a block of threshold formation, emitter repeaters and current switches. In total we receive 38 elements.

The proposed ternary half-adder is built on the basis of only one multi-threshold element of multi-valued logic, which allows the use of only 11 elements. That is, the proposed scheme contains more than 3 times less logical elements.

Disjunction and strong conjunction

On the basis of MTEML the structures of elements of realization of some ternary functions, namely - disjunctions and strong conjunctions were constructed.

A two-threshold MTEML is sufficient to construct a disjunction, while a four-threshold MTEML is required to construct strong conjunctions.

The values of the outputs (truth tables) of these functions are given in table 6 and table 7, and block diagrams in Fig. 7 and fig. 8 respectively.

The tables show which current switchers outputs need to be combined in block diagrams to obtain disjunction and strong conjunction functions.

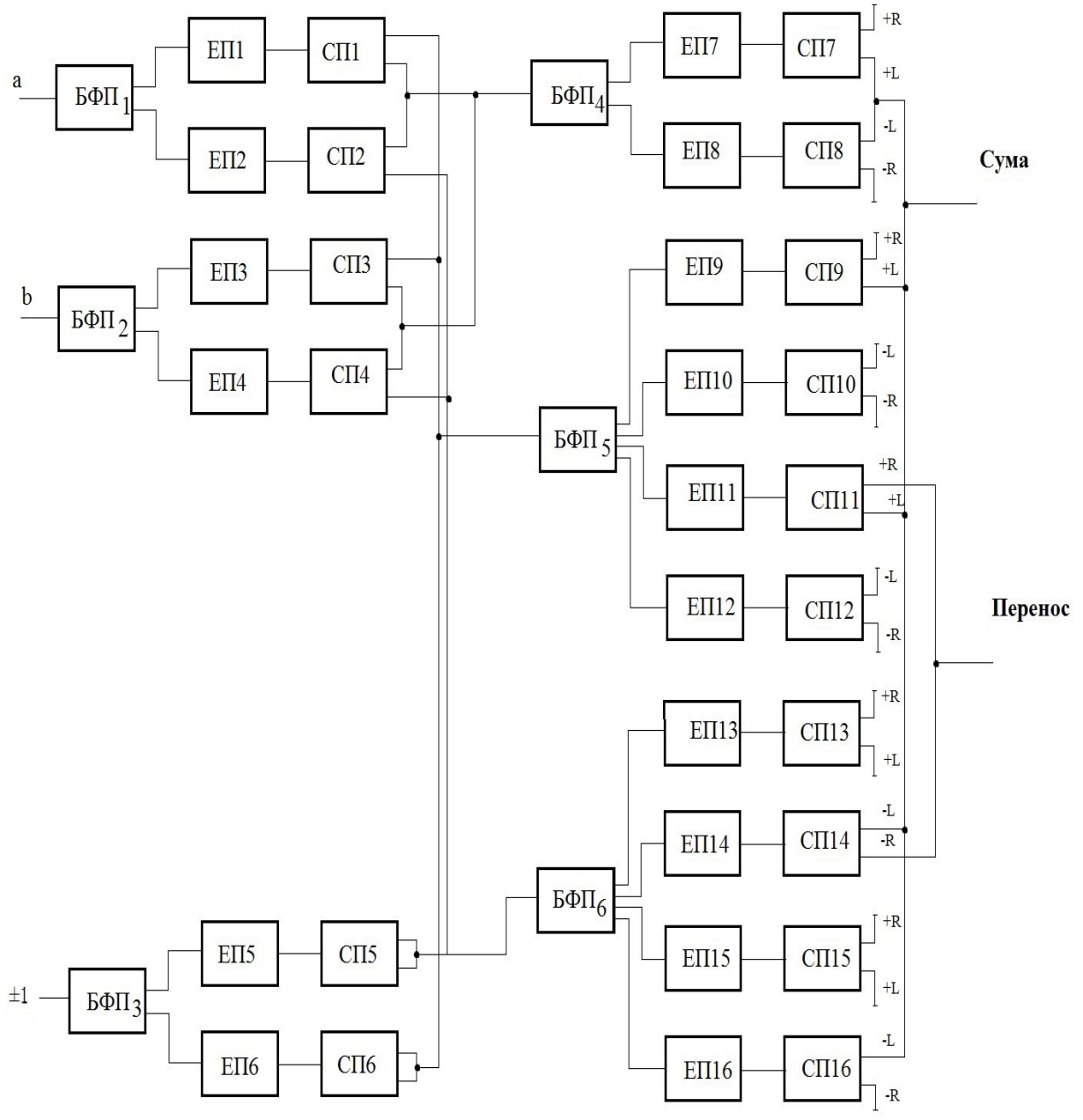


Fig. 6: The structure of the ternary half-adder on the basis of TETL, reduced to a form similar to the half-adder built on the basis of MTEML

Table 6

Table for constructing a ternary disjunction

a	b	terlev	$a \vee b$	$+\mathbf{R_1}(ab)$	$-\mathbf{R_1}(ab)$
–	–	– –	–	0	–
–	0	–	–	0	–
–	+	0	0	+	–
0	–	–	–	0	–
0	0	0	0	+	–
0	+	+	+	+	0
+	–	0	0	+	–
+	0	+	+	+	0
+	+	++	+	+	0

Table 6 shows us, that to obtain the disjunction a∨b it is necessary to combine the outputs $+R_1$ та $-R_1$, , ie we can write:

$$a \vee b = +R_1(ab);\ -R_1(ab), \text{or reduced: } a \vee b\ =\ +R_1;\ -R_1$$

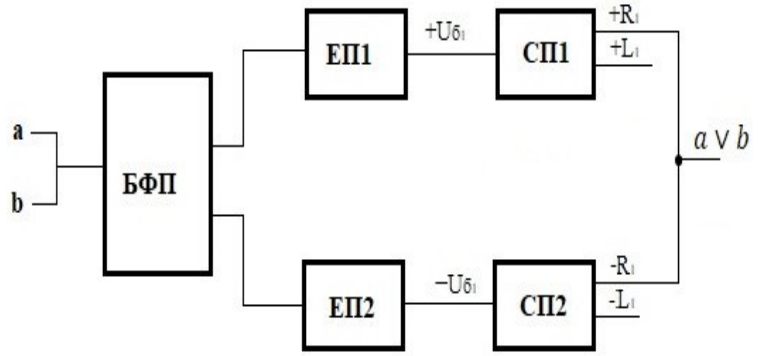


Fig. 7: Ternary disjunction

Table 7

Strong conjunction table

a	b	terlev	$a\&_L b$	$-\mathbf{R_2}(ab)$	$+\mathbf{R_1}(ab)$	$-\mathbf{R_1}(ab)$	$+\mathbf{L_1}(ab)$
–	–	– –	–	–	0	–	+
–	0	–	–	–	0	–	+
–	+	0	–	–	+	–	0
0	–	–	–	–	0	–	+
0	0	0	–	–	+	–	0
0	+	+	0	–	+	0	0
+	–	0	–	–	+	–	0
+	0	+	0	–	+	0	0
+	+	++	+	0	+	0	0

To obtain a strong conjunction a&b it is necessary to combine the outputs $-R_2$, $+R_1$, $-R_1$, $+L_1$, i.e. we can write

$$a\&b = -R_2;\ +R_1;\ -R_1;\ \ +L_1$$

In this case, we see that the signal -R2 is used to obtain a strong conjunction, which is fundamentally absent in the TETL.

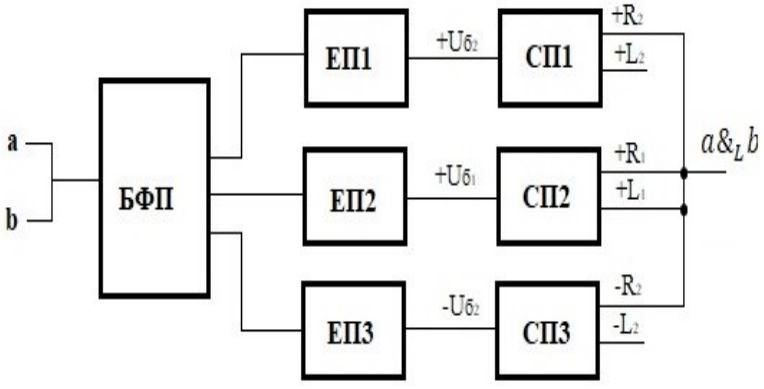


Fig. 8: Block diagram of a strong conjunction

Conclusions

Thus, the threshold element of ternary logic on bipolar unsaturated transistors and such elements of ternary systems based on it as the shaper of three-digit constants, ternary repeater, non-cyclic inverter, "OR" circuit are considered. All these elements do not allow to fully implement ternary logic, do not have a general approach to its implementation, and have a complex structure, which complicates the implementation of ternary devices. The multi-threshold element of multivalued logic and its specific four-threshold implementation allows to obtain ternary elements of relatively simple structure with a more generalized approach to their implementation. The use of a multithreshold element of multivalued logic as a basis for the construction of elements of ternary systems is proposed, the advantages of its use are determined. The realization of half-adder, disjunction and strong conjunction is given.

The implementation of ternary devices based on threshold logic is a way to create ternary devices that can compete with binary devices in terms of equipment, capacity, operational capabilities, and variety. This, in turn, leads to greater speed and simplification of the structure of devices, as well as increasing the speed of data processing. These advantages are significant in such areas of computer use as intelligent data processing systems, expert systems, decision theory, i.e., where data analysis is performed.

Acknowledgements

The work was supported by a research topic №306 Computer technologies, systems, components: modeling and software (№ state registration 0121U110545).

References

1. D. Michael Miller; Mitchell A. Thornton (2008). Multiple valued logic: concepts and representations. Synthesis lectures on digital circuits and systems 12. Morgan & Claypool Publishers. ISBN 9781598291902.
2. Yu. Gunchenko, L. Martynovych, K. Tikhonova, D. Slutsky, The concept of construction of devices for ternary logic - Proceedings of the V All-Ukrainian scientific-practical conference Perspective directions of modern electronics, information and computer systems (MEICS-2020) Dnipro , November 25–27, 2020 - pp.93-94.
3. Levchuk VV, Gunchenko YO, Emelyanov PS, Elements based on ternary logic // Abstracts of the Fourteenth All-Ukrainian Conference of Students and Young Scientists "Informatics, Information Systems and Technologies", Odessa, April 14 2017 - Odessa 2017. - p.112-114.
4. Pat. 2394366 Russia, IPC (2006.01) H03K19 / 00. Threshold element of ternary logic and elements based on it. Publ. 10.07.2010.
5. Pat. 2461122 Russia, IPC (2006.01) H03K19 / 00. Node of ternary circuitry and decoders-switches based on it. Publ. 10.09.2012
6. Pat. 2015130589 Russia, IPC (2006.01) H03K19 / 00. Triple reverse shift register. Publ. 27.05.2016
7. Pat. 1563821 United Kingdom. Ternary logic circuits with CMOS Devices / Hussein Talaat Mouftah. March 19, 1976, publ. 02.04.1980.
8. Pat. UA 118735 Ukraine, IPC (2017.01) H03K19 / 00. Multitreshold element of multivalued logic / Gunchenko Yu.O. Application 23.03.2017, publ. 28.08.2017.
9. Gunchenko Yu., Levchuk V., Kuznichenko S., Oleynyk O., The concept of Construction of Logical and Arithmetic Devices for Multivalued Logics // International scientific-practical. conf. "Information technologies and computer modeling", Ivano-Frankivsk, 2018. - P.216.
10. Levchuk VV, Gunchenko YO, Emelyanov PS, Four threshold element of ternary logic // II All-Ukrainian scientific-practical conference of students, graduate students and young scientists "Information technologies in modeling", Mykolaiv -March 24, 2017 - Mykolaiv 2017. - p.140.
11. Pat. UA 122996 Ternary half-adder based on a multi-threshold element of multivalued logic / Gunchenko YO, Lenkov SV, Malakhov EV, Shvorov SA, Ustymchuk VV, Lukin VE, Mezhuyev VI, Lenkov ES, Levchuk V. V., Application. 16.06.2017, publ. 12.02.2018.
12. Levchuk V., Gunchenko S., Ternary Half-Adder // 55 Confrence of Stundent's Scientific Cirles, Krakow, 10 may 2018. – c. 264.

Larysa Martynovych **Лариса Мартинович**	Post-Graduate student of Computer System and Technology Department, Odesa I.I. Mechnykov National University, Odesa, Ukraine, e-mail: larysa.yaroslavna@gmail.com, https://orcid.org/0000-0001-7351-1467	аспірантка кафедри комп'ютерних систем та технологій, Одеський національний університет імені І.І. Мечникова, Одеса, Україна.
Yurii Gunchenko Юрій Гунченко	DrS, Professor, Head of Computer Systems and Technology Department, Odesa I.I. Mechnykov National University, Odesa, Ukraine, e-mail: gunchenko@onu.edu.ua. https://orcid.org/0000-0003-4423-8267	доктор технічних наук, професор, зав. кафедри комп'ютерних систем та технологій, Одеський національний університет імені І.І. Мечникова, Одеса, Україна.
Yurii Shugailo **Юрій Шугайло**	c.p.- m.s., as.prof., Department of Computer Systems and Technology, Odesa I.I.Mechnykov National University, Odesa, Ukraine, e-mail: y.shugailo@gmail.com. https://orcid.org/0000-0003-2144-093X	канд фіз-мат. наук, доцент кафедри комп'ютерних систем та технологій Одеського національного університету імені І.І. Мечникова, Одеса, Україна.
Yurii Bercov **Юрій Берков**	as.prof., Department of Computer Systems and Technology, Odesa I.I.Mechnykov National University, Odesa, Ukraine e-mail: bercov@onu.edu.ua https://orcid.org/0000-0001-9704-1121	ст. викладач кафедри комп'ютерних систем та технологій Одеського національного університету імені І.І. Мечникова, Одеса, Україна.

UDC 519.6:94
https://doi.org/10.31891/csit-2022-4-9

OLEKSANDR MELNYKOV, ANDREI KAPELESHCHUK
Donbas State Engineering Academy, Kramatorsk, Ukraine

WEB-BASED SYSTEM of DECISION SUPPORT FOR CALCULATING COMBAT AND NON-COMBAT LOSSES DURING MILITARY CAMPAIGNS IN THE MIDDLE AGES

The project is devoted to calculating the number of losses of medieval armies using artificial intelligence methods. Several possible calculation methods have been considered, and historical sources have been analyzed. The analysis of factors that affected combat and non-combat losses and the analysis of scientific and popular science literature have been carried out, and the methods proposed by specialists have been translated into the language of formulas. The process of building neural networks, selecting their architecture, searching and synthesizing data for training samples have been considered, and the process of training and verifying the obtained data have been considered. Two models of neural networks and an information model in the form of UML diagrams of the future web application have been developed. Diagrams of use cases, classes, and components for each element of the web system have been described. The models have been implemented using the modern Django framework. A full-fledged web application with microblogging has been developed and tested. The conclusions have been drawn about the efficiency and scalability of the developed system, and the functionality of the system has been demonstrated using a real historical example.

The project has developed the models of an information system for analyzing the number of troops and calculating military losses of the remote past. Mathematical models have been described, historical sources have been analyzed, and detailed models of the application have been created using the UML modeling language, which allows you to understand its interface in detail at the modeling stage.

Based on the data obtained in the analysis of historical literature, the neural network architectures have been developed to determine non-combat losses in the medieval army and determine combat losses based on the data on the number of each branch's soldiers of the armed forces who were involved during the battle.

Web applications for calculating combat and non-combat losses have been developed, interface design and mini-block for publishing system News have been developed either. The non-combat losses suffered by the Mongol army in the first month of the Western campaign (1236) have been calculated as the example.

Keywords: medieval military history, calculating the number of troops, Perceptron, modeling, Python, Django.

ОЛЕКСАНДР МЕЛЬНИКОВ, АНДРІЙ КАПЕЛЕЩУК
Донбаська державна машинобудівна академія

WEB-СИСТЕМА ПІДТРИМКИ ПРИЙНЯТТЯ РІШЕНЬ ДЛЯ ПІДРАХУНКУ БОЙОВИХ ТА НЕ БОЙОВИХ ВТРАТ ПІД ЧАС ВОЄННИХ ПОХОДІВ У СЕРЕДНЬОВІЧЧЯ

Робота присвячена розрахунку чисельності втрат середньовічних армій за допомогою методів штучного інтелекту. Розглянуто декілька можливих методів підрахунку, проаналізовано історичні джерела. Проведено аналіз факторів, що впливали на бойові та не бойові втрати, аналіз наукової та науково-популярної літератури та перекладено запропоновані спеціалістами методи на мову формул. Розглянуто процес побудови нейронних мереж, підбору їх архітектури, пошуку та синтезу даних для навчальних вибірок, розглянуто процес навчання та перевірки отриманих даних. Розроблено дві моделі нейронних мереж та інформаційна модель у вигляді UML-діаграм майбутнього веб-додатку. Описано діаграми варіантів використання, класів та компонентів для кожного елементу web-системи. Виконано реалізацію моделей за допомогою сучасного фреймворку Django. Розроблено та опротестовано повноцінний web-додаток з мікроблогом. Зроблено висновки щодо ефективності та можливості масштабування розробленої системи, продемонстровано функціонал системи на реальному історичному прикладі.

У рамках проекту розроблено моделі інформаційної системи для аналізу чисельності та підрахунку військових втрат далекого минулого. Описано математичні моделі, проаналізовано історичні джерела та створено детальні моделі додатку за допомогою мови моделювання UML, що дозволяє детально зрозуміти його інтерфейс на етапі моделювання.

На основі даних, отриманих під час аналізу історичної літератури, розроблено архітектури нейронної мережі для визначення небойових втрат середньовічної армії та визначення бойових втрат на основі даних про кількість солдатів кожного з видів збройних сил, які були загиблі. брав участь під час бою.

Розроблено веб-додатки для розрахунку бойових і небойових втрат, розроблено дизайн інтерфейсу та міні-блок для публікації системи Новини. Для прикладу підраховано небойові втрати монгольської армії в перший місяць західного походу (1236 р.).

Ключові слова: середньовічна військова історія, розрахунок чисельності військ, персептрон, моделювання, Python, Dgango.

Introduction

War is a phenomenon that accompanies humanity in all historical time, so the question of the number of troops and military losses was also raised by Herodotus. Unfortunately, there is currently no single system for calculating combat losses, and even losses in fairly modern conflicts are mostly estimated approximately, and during the constant wars of the Middle Ages, there was simply no one to calculate the military losses. It should be noted that by the end of the XIX century, up to 70% of losses fell on the so-called non-combat losses (death from diseases, injuries, cold, etc.).

When it comes to estimating losses, first you need to estimate the size of armies, because the number of troops directly affects losses, but the number of armies directly depends on losses. During the three centuries of the development of modern historical science, historians have developed several approaches to estimating the size of the medieval army:

The first and oldest approach was to directly reading historical sources and trusting them. But in the chronicles, either there is no information about the number of troops at all, or their number was determined in an unknown way (it is not clear who counted the soldiers, considering that the chronicles, as a rule, were created several decades after the events described in them). The Chronicles rather show the chronicler's attitude to certain events: for example, in the European medieval chronicles, there were always thousands of times more dead pagans than Christians, and it may seem that the Invincible Knight's army came out of the battle without any losses, but this is not like that.

The second approach is based on calculating the number of troops by determining the number of soldiers in one combat unit, and then determining the number of these combat units. Many historians use this method, but it does not take into account the division into combat and non-combat losses and assumes that each "thousand" has a thousand soldiers, and not less or more soldiers. But even in any modern army, there is no fully equipped military unit (even in those that do not fight), not to mention the colossal non-combat losses during the marches.

Another approach involves calculating the number of troops, which is based on the total number of inhabitants of the country and on the size of the enemy army (of course, no one takes more troops than it is necessary). Based on the archaeological data (the area of cities, the number of towns, etc.), knowing the amount of financial resources needed for the birth of one warrior (especially it is typical for the Knight armies), you can calculate the mobilization reserve of the country, and relying on it you can calculate the number of troops that a particular country could send for perform combat tasks. An important role in assessing the mobilization reserve is played by the type of army that the country used. Nowadays, the mobilization reserve is almost all males of military age, and some military affairs can be taught for a few months. The model of the knight's army assumes the presence of a professional army only (chivalrous military affairs were studied for years), and the Knight's equipment required very large financial costs.

The calculated number of troops can be checked by using additional calculations of the amount of food for feeding the army, the length of the wagon train, and other factors, checking the calculated values for adequacy. For example, for each mounted warrior at different times there were two or three horses, and horses or oxen were also needed to transport wagons. According to the standards, with an average load, a medium-sized horse needs 13 feed units, that is 8 kg of high-quality hay. To maintain the mobility of the army, it is necessary to keep a reserve for several days with you, especially during winter hikes, when it is not possible to graze animals. It should be noted that during the summer period, there was also no great opportunity for grazing animals during crossings, because the grazing process takes a long time, but during military campaigns, the main task was to walk as much distance as possible in a day [1]. No more than 70 kg of hay can be loaded on one cart, wood can no longer withstand, and one cart with a horse occupies at least 10 meters on the march, a horse-drawn guesser occupies at least 6 meters. Knowing this, it is not difficult to calculate the lengths of columns, the movement of which had to be controlled without any means of communication.

The main reasons for the extremely large non-combat losses in the XIX century were: epidemics, lack of medicine, lack of field medicine, lack of a centralized food supply system (each unit supplied its own food, feeding on the site), lack of any living conditions (soldiers slept on the ground, tents were not existed). On the other side, people in the XIII century could be physically stronger and more adapted to severe weather conditions (which is confirmed by some sources) than people of the XIX century, but somehow it is impossible to measure the average strength and compare it with previous measurements, this is a rather subjective parameter. To predict non-combat losses, you can trace the main parameters that determine human health: air temperature, precipitation, and so on. Also, some data on the diseases or nutrition of people of that time can be given to us by the archaeological sources or miniatures of that time.

Air temperature is one of the most important factors, knowing which, you can calculate the number of deaths from seasonal infections, pneumonia, dehydration or frostbite. To calculate this parameter, you need to know in which climatic zone events occurred, at what time of the year and approximate years (it is known that in 1200-1300 in Europe it was much warmer than in the XX century [2]). Another equally important parameter is the terrain through which the army passes, the probability of being injured on the march or becoming a victim of poisonous or predatory animals depends on the terrain.

In the complex "time of the year-terrain-climate zone" you can describe most of the phenomena that lead to non-combat losses, but for a more accurate calculation, you also need to take into account the number of days in the campaign, since most viruses have an incubation period, which means that in the first weeks of the campaign, the intensity of this type of disease is lower.

On fig. 1 it is possible to notice the existence of some dependence of non-combat losses on the year, so we can assume the existence of a dependence on the weather conditions. (Jumps in the yellow and blue lines are caused by the plague epidemic in 1710-1711 [3, p. 58-59] – this suggests that large epidemics should be considered separately).

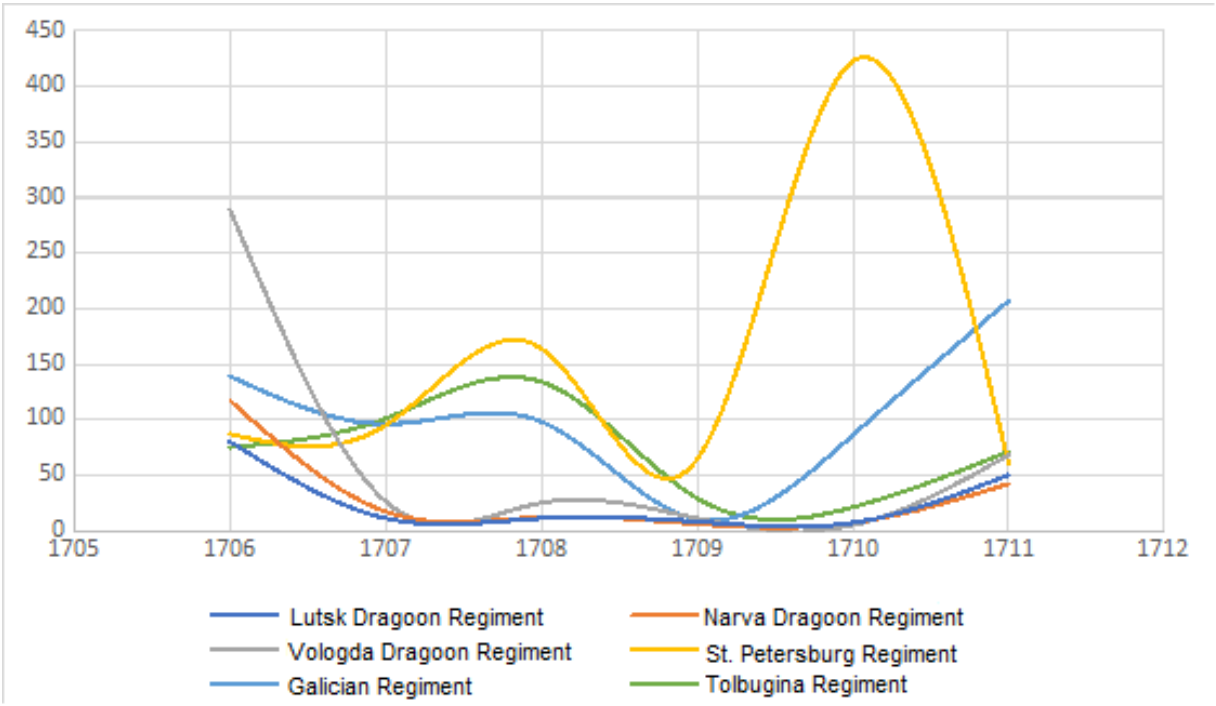


Fig. 1. Indicators of non-combat losses in various units of the Russian Empire from 1705 to 1711 [3]

Analysis of existing solutions

In 1916, the mathematician F. W. Lanchester developed a system consisted of two differential equations (1) and served to count the victims of wars:

$$\begin{cases} \frac{dx}{dt} = ax + bxy + cy + d \\ \frac{dy}{dt} = ey + fyx + gx + h \end{cases} \quad (1)$$

where:
a and *e* – speed of non-combat losses;
b and *f* – rate of losses due to impact on area targets;
c and *g* – losses from enemy actions on the front line;
d and *h* – incoming or outgoing reserves [6].

But this model is more adapted to the war of the late XIX - early XX century than for the wars of the Middle Ages, because such parameters as *b* and *f* are absent during medieval clashes, and primitive non-firearm artillery was used mainly for storming and breaking through the defenses of fortresses, rather than for destroying enemy manpower in an open field, for medieval battles the linear model of Osipov-Lanchester is more suitable.

Another approach to modeling considers combat in the form of a computer game with different levels of detailing, depending on the power of the computer (both an entire regiment and one warrior can be taken per unit, depending on the available computing resources). Most variants of such generalized models do not take into account the peculiarities of climate and terrain [7]. Some of the game projects of the Wargaming studio can be used to simulate the fighting of the XX century, due to the fact that they take into account many real physical indicators, modeling medieval battles with the help of games can provide some new information about the events. However, such modeling is very time-consuming, and most of the provided information will be superfluous in the task of simply counting warriors and losses.

Modeling

Let's create a mathematical model for calculating of non-combat losses. We have the following initial data:
t_{avg} – average monthly temperature;
T – deviation from the climatic optimum;
d – number of days on the hike;
k_t – terrain difficulty factor;
kol – number of warriors participating in the campaign;
os – average monthly precipitation.
a – non-combat losses per day.

The number of deaths from diseases can be calculated using formulas 3 and 4.

$$t = t_{cp} + T \tag{2}$$

$$c_1 = \begin{cases} \text{const1}, \text{at t} < 5 \\ \text{const2}, \text{ at } 5 < \text{t} < 15 \\ \text{const3}, \text{ at t} > 15 \end{cases} \tag{3}$$

$$d = consv + c_1 \tag{4}$$

Constants are the number of deaths as a percentage, determined by selection, based on the data from the wars of the XIX century. Constants 1-3 correspond to various climatic diseases (seasonal flu, hypothermia, cold, pneumonia, etc.), but constant v is responsible for diseases that do not depend on the season of the year.

Due to the fact that in most regions of Europe the amount of precipitation is quite small in summer, the coefficient is the lowest in summer, and in the winter months a good layer of snow, on the one hand, makes it difficult to move, on the other hand – places like swamps become less dangerous. The most dangerous months are March, April, October and November, when there is quite a lot of cold precipitation and the air temperature decreases. You can see that the "danger" increases in proportion to how much the air temperature decreases, and the amount of precipitation increases. For the winter months, the" danger " should be reduced. Let's write all mentioned above as Formula (5):

$$p = \frac{1}{t} * os - zile \tag{5}$$

where zile – the coefficient of cold winters and hot climate – it takes the value 0 at a comfortable plus temperature; at subzero or extremely plus temperatures – the value greater than 0.

Total non-combat losses per day:

$$a = (p + d)\frac{kol}{100} \tag{6}$$

Formula (6) does not take into account large epidemics, but it takes into account the usual set of viral and bacterial diseases. To calculate epidemics, we should add the percentage of deaths during the epidemic to the numbers *p* and *d*.

There is only the one uncertainty left in our model, namely the value of all our constants. They must be determined separately, calculating the result each time. Therefore, it is appropriate to use a direct propagation neural network – a perceptron.

At the perceptron inputs, we provide the following parameters: average temperature per month, average precipitation per month, terrain number, and the number of days on a hike. So on the first layer, we have got 4 neurons. We have 6 constants, the values of which need to be selected, so on the second layer we place 6 neurons, on the third layer – one output neuron, which gives out the percentage of non-combat losses per day. The neural network model is shown in fig. 2. Perceptron gives out the percentage of deaths per day.

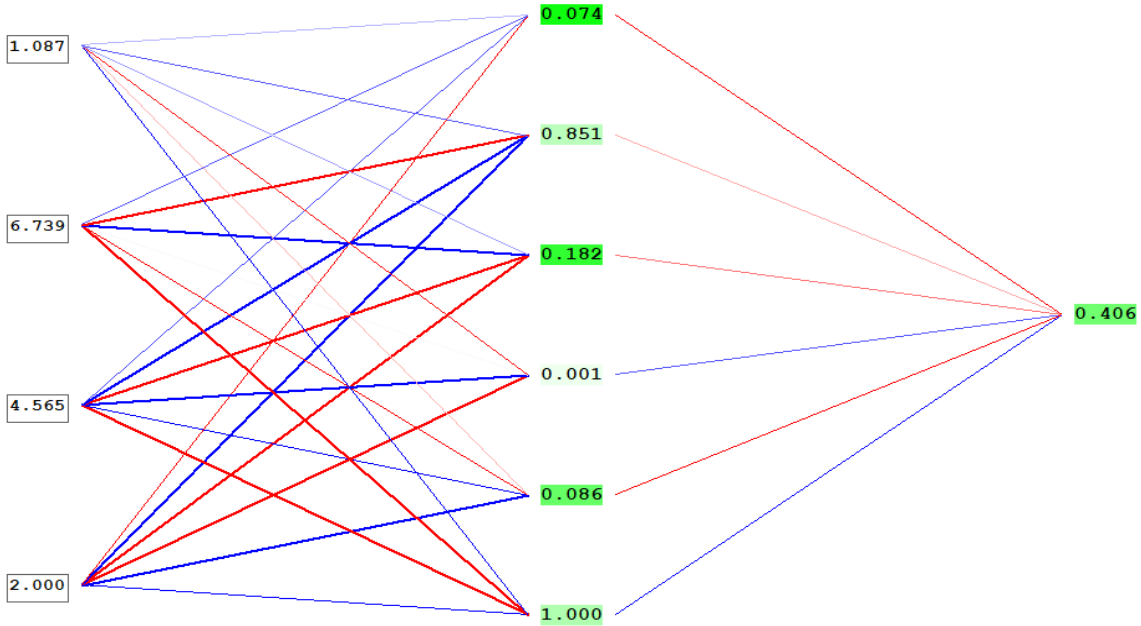


Fig. 2. Neural network model in the NeuroNet environment

But we take the average daily temperature and precipitation from fairly unverified sources (it is impossible to find out the amount of precipitation in the Middle Ages), so we propose a different neural network model that takes into account only the type of terrain, the time of the year, and the number of days in the hike (Fig. 3). Instead of temperature and precipitation data, we add two inner layers (which emulate temperature and precipitation) with 12 neurons, according to the number of months in the year.

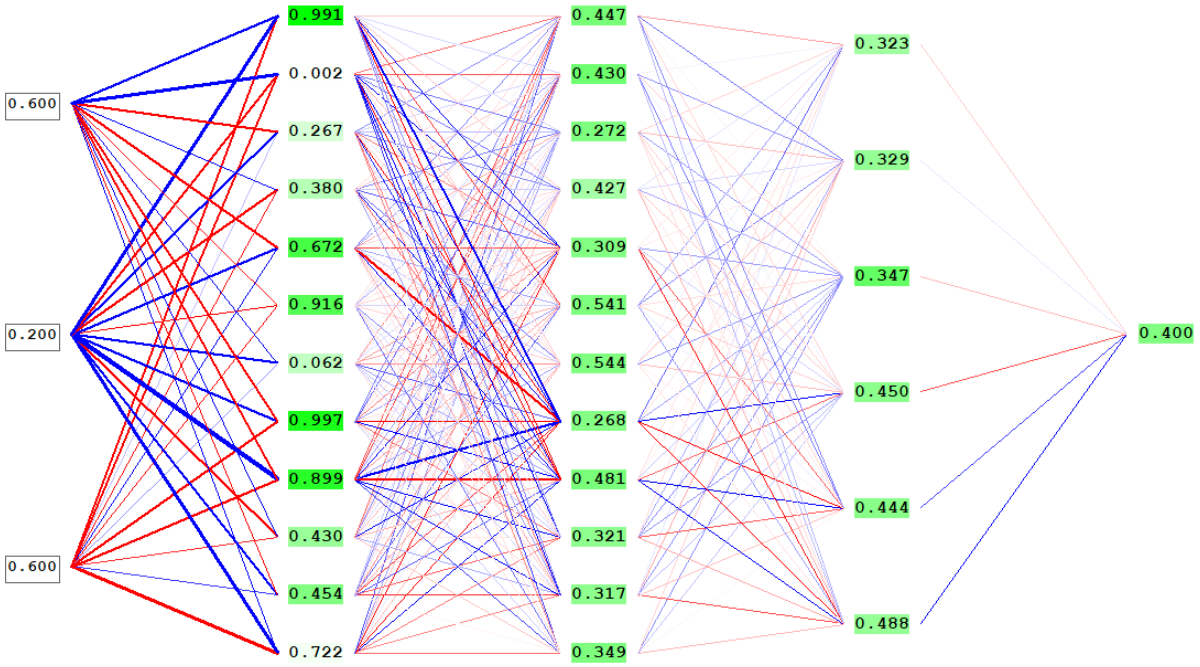


Fig. 3. Neural network model in the NeuroNet environment

To create test models, the training data was created in such a way that it was very similar to real ones (the real temperature and precipitation indicators were used). The goal of this stage is to determine the very possibility of such calculations and choose the architecture of the neural network. To submit the data to the network, they are normalized in the range from 0 to 10.

At the stage of final training, the static data on non-combat losses of the XVIII-XIX centuries is used to clarify the weighting coefficients.

Next, we calculate the number of losses during combat operations. In the Middle Ages, there were two types of such actions: military clashes and siege fortresses. To calculate the former, it is possible to apply the linear Osipov – Lanchester's law [6]:

$$A_0 - A_t = E(B_0 - B_t) \tag{7}$$

where:

A_0 – primary number of units of side A;

A_t – the number of troops remaining in army A at the time of T;

B_0 – primary number of units of Side B;

B_t – number of troops remaining in army B at the time of T;

E – weapon quality ('E'xchange Rate) = (striking ability of side B's weapon) + (striking ability of side A's weapon), (striking strength) = (weapon Quality Factor) * (number of units) [4].

However, such a model cannot work in the case of sedimentary military operations or when different branches of the armed forces collide, since it provides for the principle "one soldier kills one other soldier". To calculate the losses in the "heavy horse cavalry – light infantry" collision, the losses among the infantry are significantly higher than among the cavalry, and it is almost impossible to calculate the losses during the storming of fortresses, since there are several options for storming the fortress to die. Namely, the death of soldiers in clashes at the walls of the fortress, the death from starvation in a fortress that is under siege, the death during the storming of the fortress, and so on.

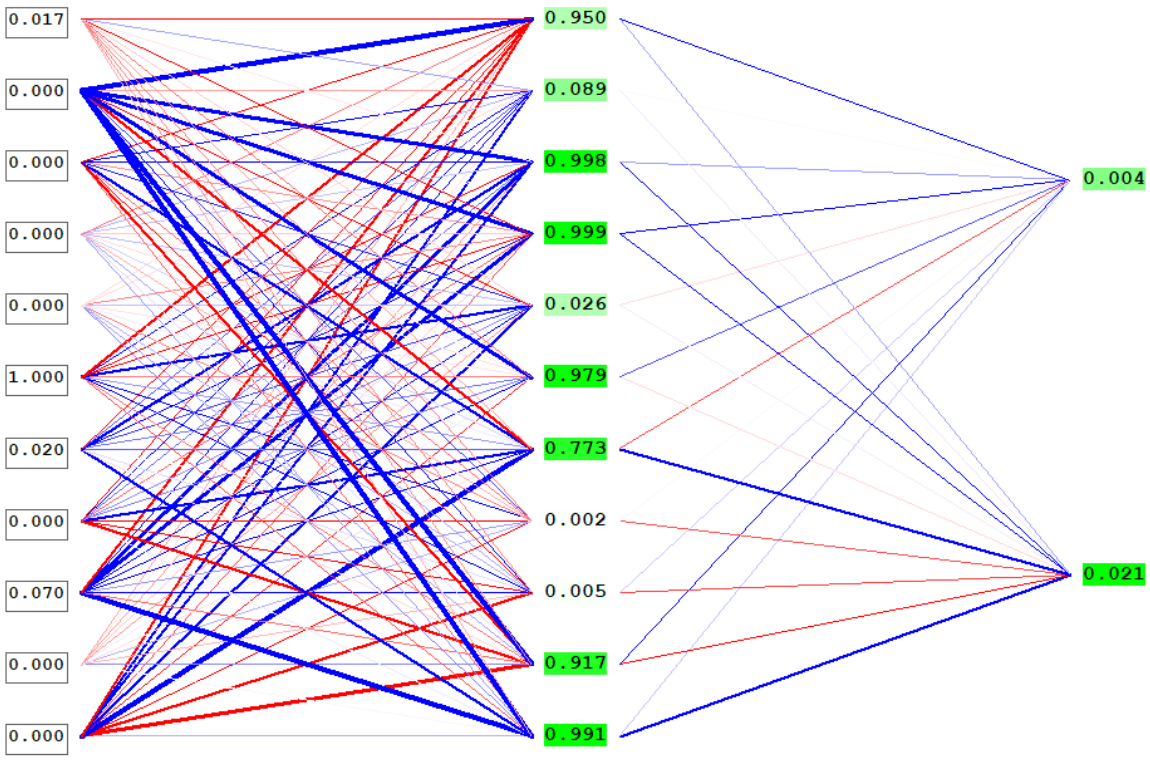


Fig. 4. Neural network model, in the NeuroNet environment

Because of this, it is proposed a less accurate, but a more flexible model for approximate counting of victims of those collisions, the number of victims of which, even approximately, was impossible to estimate using classical methods. Knowing which forces participated in a combat collision, you can calculate approximate combat losses based on the data on losses in similar collisions. Thus, after analyzing a large number of battles, you can deduce some patterns. For such purposes, a direct distribution neural network can be better suited. To clarify the number of inputs, we list all branches of the armed forces: heavy cavalry, light cavalry, infantry, archers, non-professional soldiers (people's rebels, militias, etc.). The result of the battle will be submitted for one more input (1 – the first army won, -1 – the second army won, 0 – the battle ended without any significant results on both sides). Thus, the network has 11 inputs and 2 outputs, and another hidden layer of 11 neurons is also needed (Fig. 4). To train the network, a small training sample was created from the real data taken from open sources. At the outputs, we get the number of losses for each of the conflict sides.

Let's estimate the size of the medieval army. The main criteria by which you can assess the correctness of the assumption for the probable number of troops are:

- column length on the march (S);
- length of the wagon train (S_{wagon});
- main column length (S_{column}).

$$S = S_{wagon} + S_{column} \tag{8}$$

As noted earlier, for normal movement on the march, a horse and a rider needed 6 meters, then (as a rule, the number of rows is taken no more than 2, because the roads in the Middle Ages were quite narrow):

$$S_{column} = \frac{kol \cdot 6}{r} \tag{9}$$

where:

r – number of rows during the transition.

To feed the army per day, you will need m_1 ton of food (based on the calculation that one adult man needs 1 kg of food per day):

$$m_1 = \frac{kol + \sum kol \;\; i_j \;\; v_j \;\; p_j}{1000} \tag{10}$$

where:

i – the number of horses of breed *j*;

v – the weight of hay that one horse of breed *j* needs;

p – the percentage of breed *j* horses in the army.

Thus, knowing the required mass of forage, you can calculate the length of the train using the following formula:

$$S_{Swagon} = \frac{m_1 - a \cdot kol + m_2}{0{,}07} \cdot 8 \quad (11)$$

where:

m_2 – other non-food cargo: siege vehicles, trebuchets, wounded, prisoners, engineering troops.

a – the average weight of the load that is on the horse with the rider.

In this way, you can calculate the length of the column on the march, as well as the time it takes for the rider to travel from the head of the column to the end or vice versa. If the rider cannot cover this distance in a day, the column can be considered unmanageable, that is, it could not exist in reality.

Creating a Web-system

Based on the proposed mathematical models, a WEB system was developed with applications for calculating combat and non-combat losses. The system provides several use cases, namely (Fig. 5):

– calculation of non-combat losses;

– calculation of combat losses;

– total loss calculation;

– checking the number of troops.

Calculating non-combat losses and checking the number involves launching the first browser application, while the task of calculating combat losses is to launch the second browser application. The task of calculating total military losses involves running both applications.

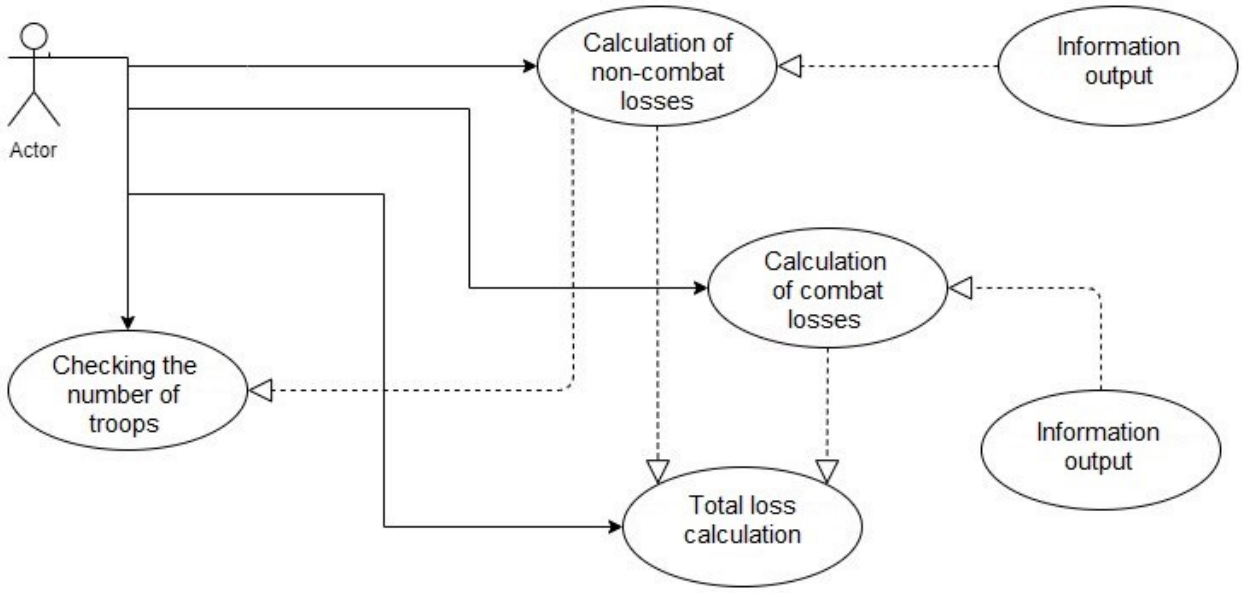


Fig. 5. Diagram of Information System use cases

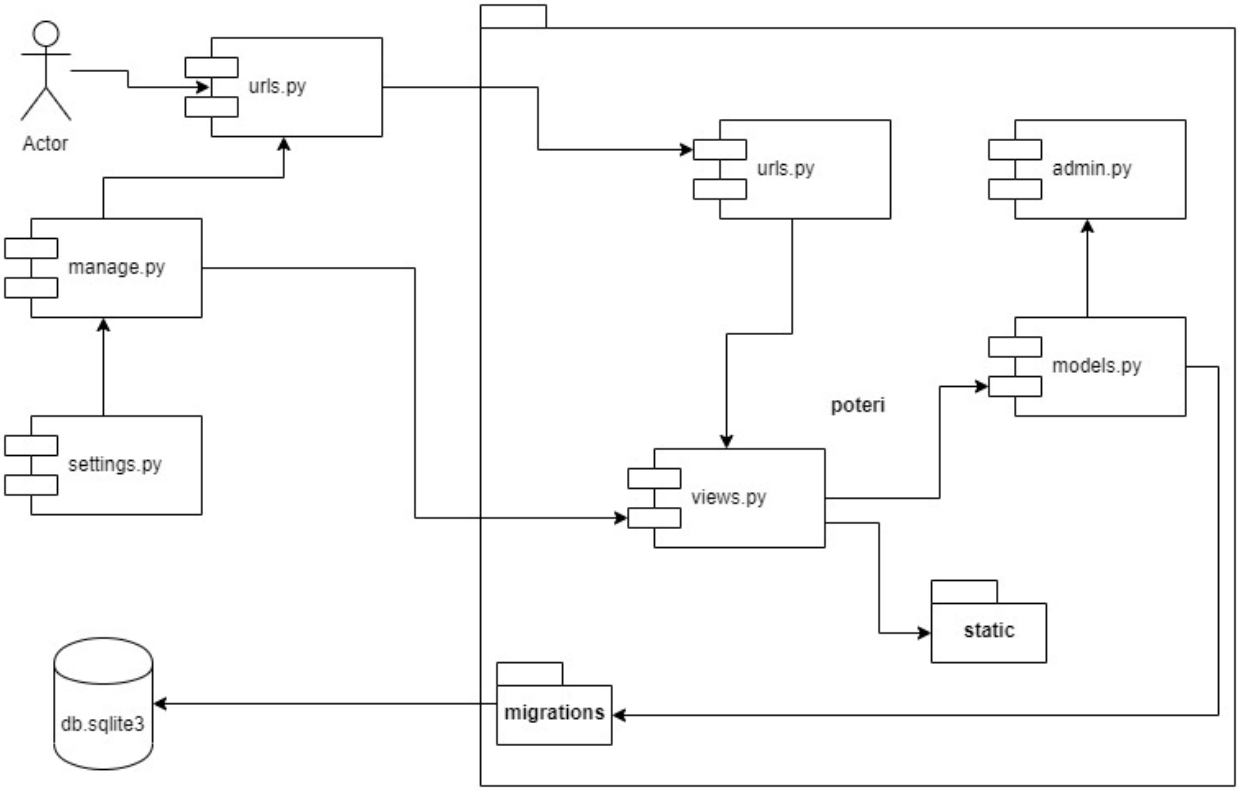


Fig. 6. Component diagram – Django file and folder structure

The Django framework was chosen as the basis for developing the application, Python was chosen for writing browser logic, and the Brython library was chosen as the translator. In the file urls.py the binding URL is located, i.e. they connect the link URL and the function that is called out when requesting this link. The functions themselves are located in the file views.py, they already access HTML templates which are stored in the folder settings.py (Fig. 6).

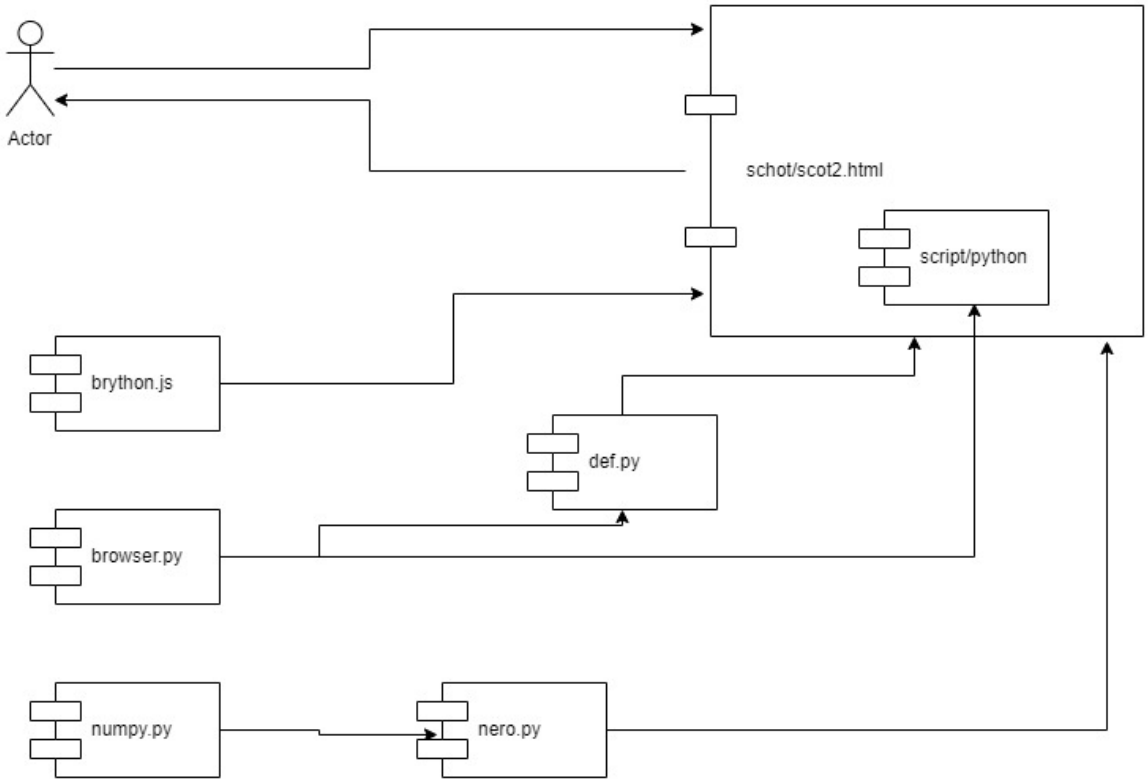


Fig. 7. Component diagram – structure of a browser application for calculating non-combat losses

The code embedded in the HTML file accesses the browser library and also calls out functions from files def.py, where the basic logic of the application is stored, and file nero.py, where the neural network will be located. The Brython file.js is a translator of Python code to JS, browser.py files and numpy.py are Python libraries for managing page elements and for mathematical calculations that will be needed when creating a neural network (Fig. 8, Fig. 9).

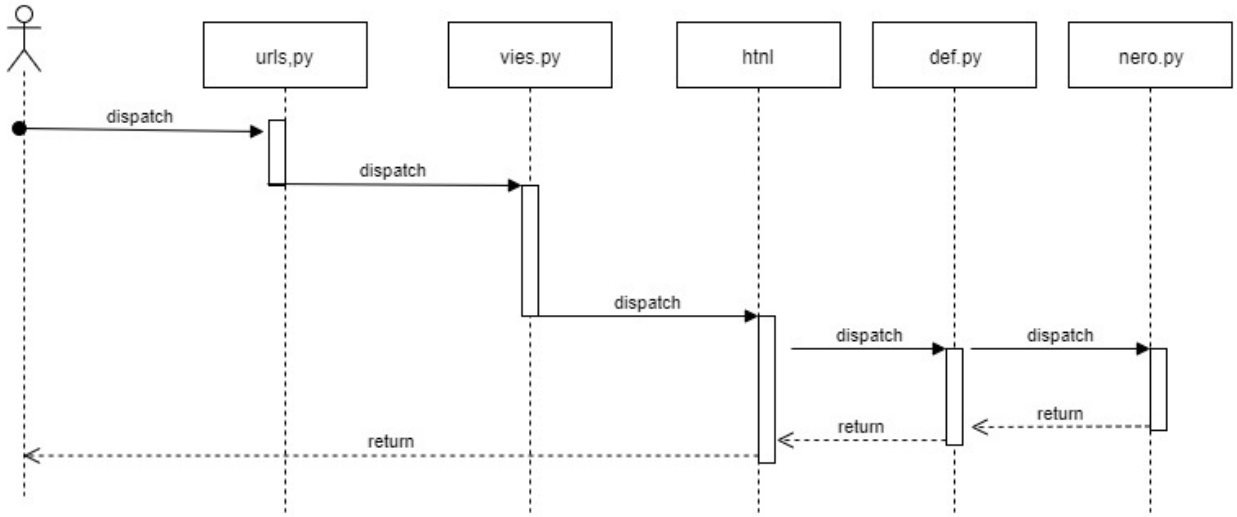


Fig. 8. Sequence diagram, non-combat loss counting process

A special feature of the Brython translator is that all files with code are connected not to each other, but to the main HTML file. Libraries that are embedded in the translator are connected in the usual way, but the modules work as shown in the class diagram (Fig. 9).

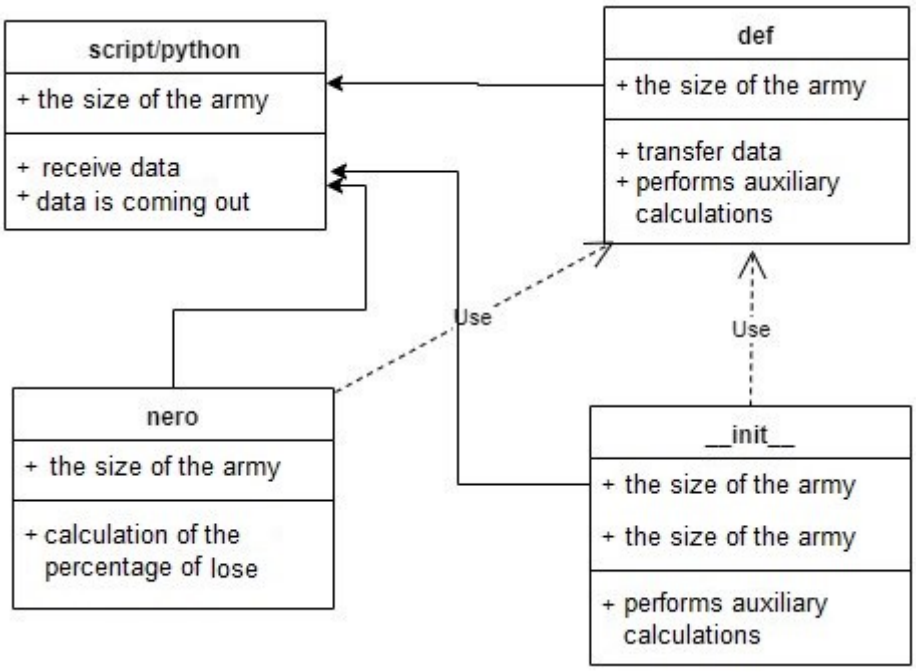


Fig. 9. Class diagram – structure of a browser application for calculating non-combat losses

The second browser application must take data from the user's fields and pass it to calculate the neural network, but in this case a static HTML page will be used (Fig. 10), the files have the same purpose, but this application does not have a script embedded in the HTML file itself.

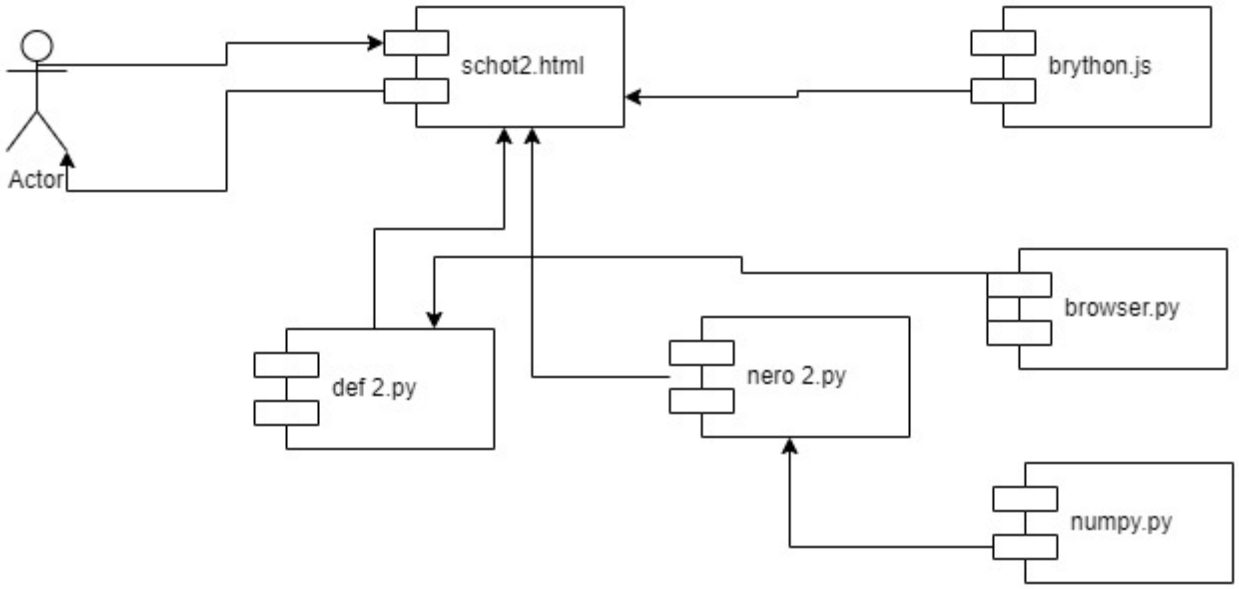


Fig. 10. Component diagram – structure of a browser application for calculating combat losses

In the final version, the system looks like a site that has the following structure: home page (Fig. 11), a project page, non-combat loss calculation tool, combat loss calculation tool, project news.

Fig. 11. Middle part of the site's home page

Since during large military campaigns the armies were divided into several columns, it would be interesting to observe the changes of all the columns at once. Before loading the home page, the user is asked to select the number of columns and the type of Army (Fig. 12).

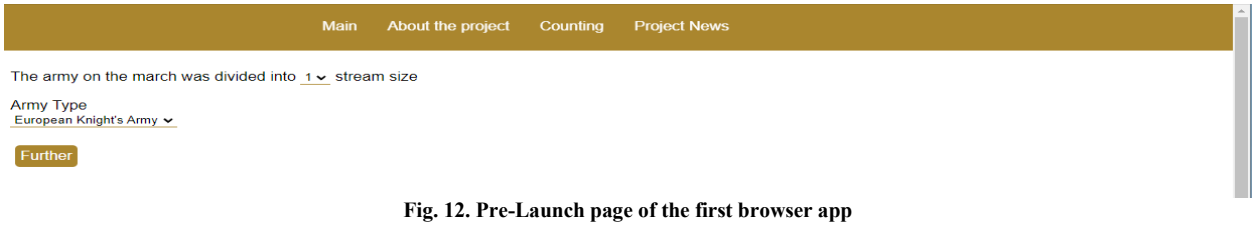


Fig. 12. Pre-Launch page of the first browser app

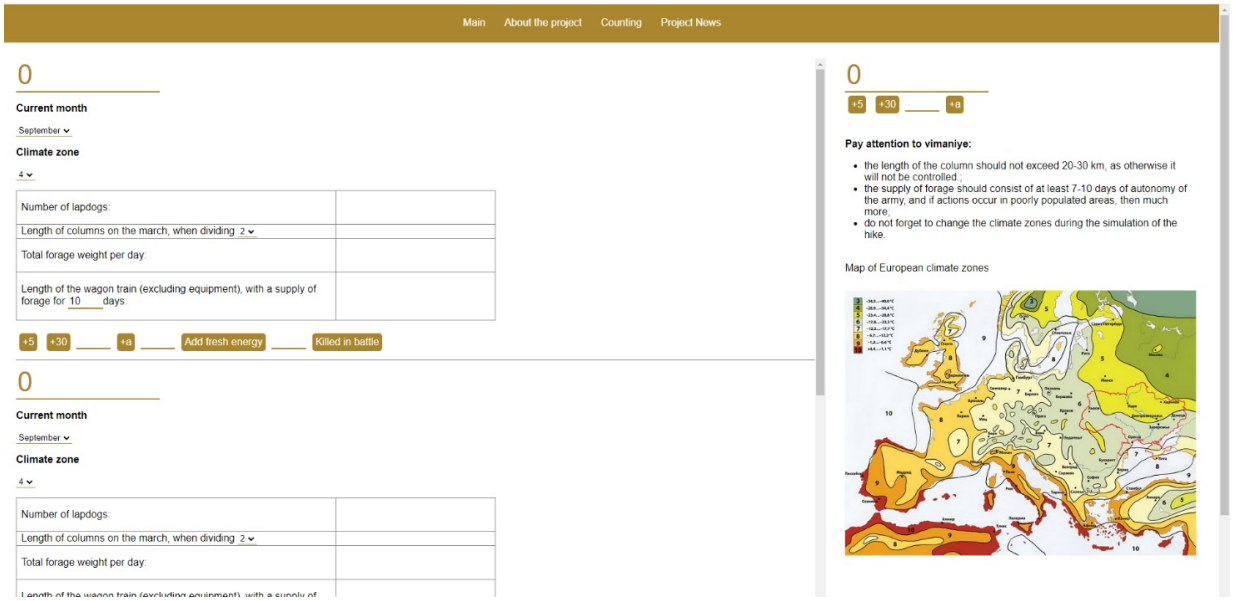


Fig. 13. Calculation of non-combat losses

To implement multiple applications simultaneously (Fig. 13) cycles are used: when rendering a page, the framework copies some sections of code a given number of times and assigns each element in each of the copies its own index, for example:

```
def func5{{ forloop.counter }}(event):
    after the rendering, it will look like this:
        def func51(event):
```

This makes it very easy to manage and create executable code from the server already at the page rendering stage. This made it possible to use each calculation block independently of the others, while leaving the possibility of unified management of all blocks. To facilitate calculations during the operation of the neural network, all weighting coefficients are already feeded in the code.

The user can add any number of campaign days, new forces, or combat losses. The current month switches automatically. the user must monitor the climate zones himself. The neural network has been trained to work only with the climatic zones in Europe.

To calculate combat losses, a neural network has also been developed that can calculate losses in a medieval combat, but it should be noted that we do not have the objective information about medieval losses, so we used loss data calculated by historians using various methods to train it. It should also be pointed out that the neural network does not take into account indirect combat losses (capture, murder of prisoners, murder of civilians, etc.) (Fig. 14)

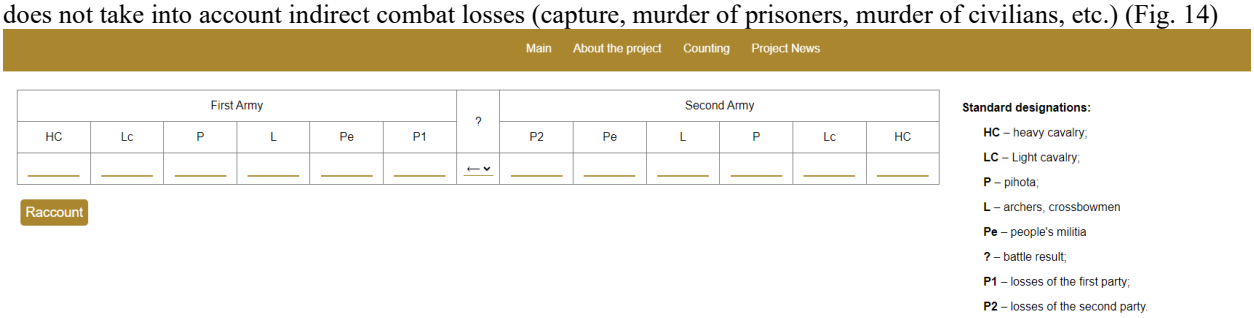


Fig. 14. Calculation of combat losses

Let's calculate the losses of the Mongol army during the eastern campaign: from historical sources, it is known that the Mongols approached our lands in the autumn of 1237, dividing into three columns. Let's find out what the number of these columns was in a month of the campaign. The total number of troops was approximately 40,000, the campaign was begun in October, the events happened near the cities of Ryazan, Suzdal and Vladimir, which corresponds to the climatic zones 4 and 5 [9, 10]. For 30 days of the campaign, non-combat losses amount to just under 3,000. The length of the largest column (15,000 soldiers from the beginning of the campaign) is 15 km, and 294 tons of food will be needed per day to feed the army. The user makes a decision about the adequacy of the column characteristics independently. The application only performs calculations based on the data fed by the user. According to the modern ideas of historians, the length of the columns could hardly exceed 20 – 30 km.

Conclusions

The project has developed the models of an information system for analyzing the number of troops and calculating military losses of the remote past. Mathematical models have been described, historical sources have been analyzed, and detailed models of the application have been created using the UML modeling language, which allows you to understand its interface in detail at the modeling stage.

Based on the data obtained in the analysis of historical literature, the neural network architectures have been developed to determine non-combat losses in the medieval army and determine combat losses based on the data on the number of each branch's soldiers of the armed forces who were involved during the battle.

Web applications for calculating combat and non-combat losses have been developed, interface design and mini-block for publishing system News have been developed either. The non-combat losses suffered by the Mongol army in the first month of the Western campaign (1236) have been calculated as the example.

References

1. Zhukov K.A. Drevniaia Rus. Ot Riuryka do Batыia / K.A. Zhukov, D.Iu. Puchkov. — S-Pb. : Pyter, 2019. – 288 s.
2. Kuvaldyn S.A. Ystoryia s klymatom. Khymyia y zhyzn 2017 - №4
3. Babych M.V. Ystoryy neboevыkh poter chastei, dyslotsyrovannыkh v Ynhermanlandyy v 1711 hodu / M.V. Babych, Y.V. Babych – Voina y oruzhye. SPb., 2012. Ch. 1. S. 47-55
4. Venttsel E.S. Osnovы teoryy boevoi эffektyvnosty y yssledovanyia operatsyi / E.S. Venttsel, Y. V. Khudiakov, Ya. M. Lykhterev, Yu. H. Mylhram – M.: VVYA, 1961 – 524 s.
5. Uvarov D. Voennыe potery v Srednevekove [Elektronnyi resurs] – URL: http://xlegio.ru/ancient-armies/military-organization-tactics-equipment/medieval-military-losses/ (15.05.2022)
6. Mytiukov N. V. Opredelenye zhertv voin cherez lanchesterskye modely/ N. V. Mytiukov //Ystorycheskaia psykholohyia y sotsyolohyia ystoryy 2009 - №2 – S. 122 – 140
7. Alekseev V. «Y bыst secha tu velyka...»/ V. Alekseev //Tekhnyka molodezhy Moskva, 1988. – №2 – S. 56–57
8. Vazhneishye bytvы Srednykh vekov. Dyletant [Elektronnyi resurs] – URL: https://diletant.media/top-5/3157/ (19.05.2022)
9. Khrapachevskyi R. Velykyi zapadnыi pokhod chynhyzydov na Bulhar, Rus y Tsentralnuiu Evropu [Elektronnyi resurs] – URL: http://xlegio.ru/ancient-armies/medieval-warfare/genghisids-great-western-campaigen/ (29.05.2022)
10. Zhaksыlyk S. O chyslennosty monholskoi armyy v Zapadnom pokhode/ S. Zhaksыlyk //Voprosы Ystoryy y arkheolohyy Zapadnoho Kazakhstana. 2010 – № 1. – S. 55 – 73
11. Kapeleshchuk A.O. Postanovka zadachi vyznachennia chyselnosti viiska u serednovichchia iz vykorystanniam matematychnoho modeliuvannia ta informatsiinykh tekhnolohii // Molodizhna nauka u konteksti suspilno-ekonomichnoho rozvytku krainy: zbirnyk tez dopovidei uchasnykiv V Yuvileinoi uchnivsko-studentskoi konferentsii, Cherkasy, 2 hrudnia 2021 r. – Cherkasy: Skhidnoievropeiskyi universytet imeni Raufa Abliazova, 2021. – S. 214-216
12. Kapeleshchuk A.O, Melnykov O.Iu. Prohnozuvannia chyselnosti viiska u serednovichchia iz vykorystanniam matematychnoho modeliuvannia // Materialy IV Vseukrainskoi naukovo-praktychnoi internet-konferentsii studentiv, aspirantiv ta molodykh vchenykh za tematykoiu «Suchasni kompiuterni systemy ta merezhi v upravlinni»: zbirka naukovykh prats / Pid redaktsiieiu H.O. Raiko. – Kherson: Vydavnytstvo FOP Vyshemyrskyi V. S., 2021. – C.228-230. – ISBN 978-617-7941-55-1 (elektronne vydannia)

Oleksandr Melnykov **Олександр Мельников**	PhD in Technics (Candidate of Technical Sciences), Associate Professor of the Department of Intelligent Decision Making Systems, Donbas State Engineering Academy, Kramatorsk, Ukraine, e-mail: alexandr@melnikov.in.ua https://orcid.org/0000-0003-2701-8051	кандидат технічних наук, доцент, доцент кафедри інтелектуальних систем прийняття рішень, Донбаська державна машинобудівна академія, Краматорськ, Україна.
Andrei Kapeleshchuk **Андрій Капелещук**	student of FAMIT, Donbas State Engineering Academy, Kramatorsk, Ukraine, e-mail: kapeleshchuk.k@gmail.com	студент спеціальності «Системний аналіз» факультету автоматизації машинобудування та інформаційних технологій, Донбаська державна машинобудівна академія, Краматорськ, Україна.

UDC 658.51.012
https://doi.org/10.31891/csit-2022-4-10

OLEH PIHNASTYI, MAKSYM SOBOL
National Technical University "Kharkiv Polytechnic Institute", Kharkiv, Ukraine

USE OF ANALYTICAL MODEL FOR SYNTHESIS OF ALGORITHMS FOR CONTROL OF TRANSPORT CONVEYOR PARAMETERS

This study presents a methodology for synthesizing optimal control algorithms for the flow parameters of a conveyor-type transport system with a variable transport delay. A multi-section transport conveyor is a complex dynamic system with a variable transport delay. The transport conveyor is an important element of the production system, used to synchronize technological operations and move material. The Analytical PiKh-model of the conveyor section was used as a model for designing a control system for flow parameters. The characteristic dimensionless parameters of the conveyor section are introduced and the similarity criteria for the conveyor sections are determined. The model of a conveyor section in a dimensionless form is used to develop a methodology for synthesizing algorithms for optimal control of the flow parameters of a transport conveyor section. The dependencies between the value of the input and output material flow of the section are determined, taking into account the initial distribution of the material along the conveyor section, variable transport delay, restrictions on the specific density of the material, and restrictions on the speed of the belt. The dependencies between the value of the input and output material flow for the case of a constant transport delay are analyzed. A technique for synthesizing algorithms for optimal belt speed control based on the PiKh-model of a conveyor section is presented. As a simplification, a two-stage belt speed control is considered. Particular attention is paid to the methodology for synthesizing optimal control algorithms based on the energy management methodology (TOU-Tariffs). The criteria of control quality are introduced and problems of optimal control of flow parameters of the transport system are formulated. Taking into account differential connections and restrictions on phase variables and admissible controls, which are typical for the conveyor section, the Pontryagin function and the adjoint system of equations are written. As examples demonstrating the design of optimal control, algorithms for optimal control of the flow parameters of the transport system are synthesized and analysis of optimal controls is performed.

Keywords: conveyor, distributed system, PDE-model, production line, belt speed

ОЛЕГ ПІГНАСТИЙ, МАКСИМ СОБОЛЬ
Національний технічний університет «Харківський політехнічний інститут», Харків, Україна

ВИКОРИСТАННЯ АНАЛІТИЧНОЇ МОДЕЛІ ДЛЯ СИНТЕЗУ АЛГОРИТМІВ КЕРУВАННЯ ПАРАМЕТРАМИ ТРАНСПОРТНОГО КОНВЕЄРА

Дане дослідження презентує методику синтезу алгоритмів оптимального керування потоковими параметрами транспортної системи конвеєрного типу із змінною транспортною затримкою. Багатосекційний транспортний конвеєр – це складна динамічна система із змінною транспортною затримкою. Транспортний конвеєр є важливим елементом виробничої системи, що використовується для синхронізації технологічних операцій та переміщення матеріалу. Як модель для проектування системи управління потоковими параметрами використана аналітична PiKh-модель секції конвеєра. Введено характерні безрозмірні параметри секції конвеєра та визначено критерії подібності до секцій конвеєра. Модель секції конвеєра у безрозмірному вигляді використана для розробки методики синтезу алгоритмів оптимального керування потоковими параметрами секції транспортного конвеєра. Визначено залежності між значенням вхідного та вихідного потоку матеріалу секції з урахуванням початкового розподілу матеріалу вздовж секції конвеєра, змінної транспортної затримки, обмежень на питому щільність матеріалу та обмежень на швидкість стрічки. Проаналізовано залежність між значенням вхідного та вихідного потоку матеріалу для випадку постійної транспортної затримки. Представлено методику синтезу алгоритмів оптимального управління швидкістю стрічки, засновану на PiKh-моделі секції конвеєра. Як спрощення розглянуто двоступінчасте регулювання швидкості стрічки. Особливу увагу приділено методиці синтезу алгоритмів оптимального управління, що ґрунтується на методології енергоменеджменту (TOU-Tariffs). Введено критерії якості керування та сформульовано завдання оптимального керування потоковими параметрами транспортної системи. З урахуванням диференціальних зв'язків та обмежень на фазові змінні та допустимі керування, характерні для конвеєрної ділянки, записана функція Понтрягіна та приєднана система рівнянь. Як приклади, що демонструють проектування оптимального керування, синтезовано алгоритми оптимального керування потоковими параметрами транспортної системи та проведено аналіз оптимальних керування.

Ключові слова: конвеєр, розподілена система, PDE-модель, потокова лінія, швидкість стрічки

Introduction

Industry 4.0 is the next step in the industrial revolution. Industry 4.0 includes the following requirements for modern production: completion of full automation of digital production; control of technological parameters in real-time; use of intelligent control systems with access to the global Internet; constant interaction with external production and marketing environments. Modern industrial technologies in process control are forced to use artificial intelligence and additive forecasting methods. At the same time, operational control of production parameters using sensors and sensors is a separate problem. Particular importance is given to this problem in enterprises with a continuous production method, in which the main element of material transportation, as a rule, is a conveyor [1]. The use of modern transport systems in the production process can significantly increase the efficiency of algorithms for the operational control of the flow parameters of the production system. The conveyor acts as a link connecting production modules and uses intelligent control components and various industrial sensors to improve the efficiency of the production process, which determine the state of the parameters of the transport system [2].

When considering the concept of "Industry 4.0", conveyor transport is important in the mining industry [3]. This is determined by the following circumstances: a) firstly, due to the low unit cost, the transport conveyor is the most suitable means for transporting the material [4, 5]; b) secondly, modern transport systems consist of a large number of sections, each of which operates with certain flow parameters [6]. These parameters for sections of the conveyor system are interconnected, and contain a variable transport delay; c) thirdly, the length of the transportation route for a separate section reaches tens of kilometers and the trend towards an increase in the length of sections continues to persist [7]; e) fourthly, the specific cost of transporting material for the standard mode of operation of the transport system is 20% of the total cost of extracting the material [8]. With an increase in the number of sections and the length of the route of the transport system, the cost of transporting the material increases non-linearly and can make up the bulk of the cost of extracting the material; f) fifthly, under the standard mode of operation, the material loading factor of the conveyor section is 0.5–0.7 of the full conveyor load. Increasing the filling factor of the conveyor section with material allows for reducing transportation costs by 30-50%. However, quite often the transport system operates with a fill factor below the norm due to the use of efficient real-time control systems; f) sixthly, restrictions are imposed on the operation modes of the conveyor section, associated, for example, with the propagation of long-wave disturbances in the conveyor belt, which can cause the destruction of the conveyor belt.

Related Works

For conveyor systems, the cost of transportation is directly related to the cost of electricity that is consumed to transport the material. One of the main methods of reducing energy consumption, which is widely used in the transport of material, involves the use of conveyor belt speed control systems [10]. This method consists in increasing the linear density of the material at the input of the conveyor section by reducing the speed of the belt, which implies an increase in the material loading factor of the conveyor section. The next most effective method for reducing energy consumption is based on the use of a material flow control system from the input bunker[11]. In this case, an increase in the linear density of the material at the input of the conveyor section is achieved by increasing the input flow of material from the accumulation bunker of the conveyor section. The third method for reducing the specific energy consumption for material transportation is based on the use of energy management methodology [12]. The method is based on the use of transport system mode control systems. The operation of sections of the transport conveyor is carried out at times when the price of electricity is low. And the last significant method for reducing specific energy consumption is the method using reverse conveyors [6]. Reducing the specific cost of energy consumption occurs due to a change in the structure of the transport route as a result of a change in the direction of the flow of material in the supply section of the conveyor to the opposite direction. The control system determines the optimal route for the movement of material from several possible ones. For the synthesis of optimal control algorithms, a set of models for the section of the transport conveyor is used. The most common models of a conveyor section are FEM models of a transport system based on the finite element method [13] and DEM models of a transport system using the discrete element method [14]. These models belong to the class of numerical models and are used in flow parameter control systems that require taking into account the non-uniformity of the material flow and the variable transport delay when the material moves along the transport route. These models require significant computational resources, which imposes significant restrictions on their use in the synthesis of algorithms for optimal control of the flow parameters of the transport system. The transition from the model of a single-section transport system to the model of a multi-section transport system, consisting of several dozen sections [6], becomes a practically unsolvable problem. This is due to the almost absence of papers devoted to the design of control systems for the flow parameters of the transport conveyor, represented by numerical models. As an alternative to the class of numerical models for describing multi-section transport systems, a separate class can be proposed, which is based on models of regression equations [16] or a neural network [17]. However, to build optimal control systems using models based on regression equations or using a neural network, a sufficiently large amount of test data is required, which is not available for non-stationary modes of operation of the transport system. This circumstance is a strong limitation for the synthesis of control systems using an alternative class of models. Thus, the above analysis clearly identified the problems that make it difficult to synthesize algorithms for optimal control of the flow parameters of a modern transport conveyor: a) taking into account the uneven flow of material along the transportation route in the presence of a variable transport delay; b) a trend towards an increase in the number of sections of the transport system. The solution to these problems requires the improvement of existing models and algorithms for optimal control of the flow parameters of the transport conveyor.

Purpose

In connection with the above circumstances, taking into account the fact that the industrial Internet of things provides an opportunity to create fully automated conveyor systems, the synthesis of algorithms for optimal control of the transport system flow parameters requires not only the improvement of existing and the creation of new models of the transport conveyor, which allow us to consider the transport conveyor as an object of intelligent control but also the improvement of methods for applying models for the synthesis of algorithms for optimal control of the flow parameters of the transport conveyor. This study is devoted to demonstrating the methodology for building transport conveyor control systems based on the use of an analytical PiKh-model of the conveyor section.

Description model of a separate section of the conveyor line

The conveyor line is a kind of production line, in a one-time approximation it has the form [1]

$$\frac{\partial[\chi]_0(t,S)}{\partial t}+\frac{\partial[\chi]_1(t,S)}{\partial S}=0, \qquad [\chi]_1(t,S)=[\chi]_{1\Psi}(t,S) \tag{1}$$

under initial and boundary conditions

$$[\chi]_0(0,S)=\Psi(S) \qquad [\chi]_1(t,0)=\lambda(t) \tag{2}$$

where S_d – the length of the section in the phase space [18]); $[\chi]_0(t,S)$, $[\chi]_1(t,S)$ is the distribution density and the material flow at the moment of time t at the position, characterized by the coordinate $S\in[0,S_d]$; $\Psi(S)$ is initial distribution along the transport route; $[\chi]_{1\Psi}(t,S)$ is the normative material flow along a transport route; $\lambda(t)$ is the input material flow.

The fact that materials move at a speed that is independent of their location on the section belt allows the material flow $[\chi]_1(t,S)$ to be expressed in terms of their density $[\chi]_0(t,S)$ and the speed $a=a(t)$ of the belt. The speed of the belt can be constant or have a continuous or step regulation in time depending on the load. This made it possible to write the closed system of equations (1)–(2) in the following form [19]:

$$\frac{\partial[\chi]_0(t,S)}{\partial t}+\frac{\partial[\chi]_1(t,S)}{\partial S}=\delta(S)\lambda(t), \qquad [\chi]_1(t,S)=a(t)\cdot[\chi]_0(t,S), \qquad \int_{-\infty}^{\infty}\delta(S)dS=1, \tag{3}$$

$$[\chi]_0(0,S)=\mathrm{H}(S)\Psi(S), \qquad \mathrm{H}(S)=\begin{cases}0, & S<0,\\ 1, & S\geq 0,\end{cases} \qquad S\in[0;S_d]. \tag{4}$$

The system of equations (3)–(4) models the flow of material moving along the transportation route. The right side of equation (3) $\delta(S)\lambda(t)$ takes into account the source of the material with the intensity characterizing the line capacity $\lambda(t)$, characterizing the line capacity (ton/hour). At the initial moment of time $t=0$ there is material on the conveyor line, which is distributed along the transportation route with linear density $[\chi]_0(0,S)$. The function $\delta(S)$ determines the place where the material enters the conveyor line: $S=0$. The system of equations is closed with respect to flow parameters $[\chi]_0(t,S)$ и $[\chi]_1(t,S)$. The closure condition shows the independence of the belt speed from the place of transportation and allows you to construct a solution to the system of equations (3)–(4) with respect to the flow parameters $[\chi]_0(t,S)$ and $[\chi]_1(t,S)$. A schematic diagram of a conveyor section with an accumulating bunker at the input is shown in Fig. 1 [20]. The material flow must be supplied to the input of the conveyor line from the bunker with the intensity necessary to ensure the required specified material flow at the output.

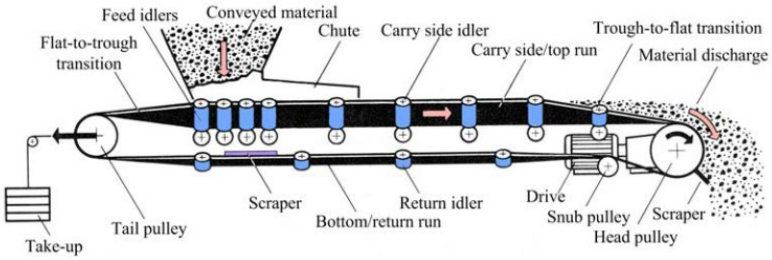


Fig.1. Schematic diagram of the conveyor line [20]

Let us supplement the system of equations (3)–(4) with an equation that determines the state of the bunker:

$$\frac{dN(t)}{dt}=\lambda_{in}(t)-\lambda(t), \qquad N(0)=N_0, \qquad 0\leq N\leq N_{\max}, \quad 0\leq\lambda(t)\leq\lambda_{\max}, \tag{5}$$

where $N(t)$ the current amount of materials in the bunker with a capacity of $N_{\max}$. The flow of materials to the input to the accumulation bunker is a given value $\lambda_{in}(t)$. Let us represent the system of equations (3)–(5) in a dimensionless form and will use the dimensionless parameters [21]:

$$\tau = \frac{t}{T_d}, \qquad \xi = \frac{S}{S_d}, \tag{6}$$

$$\theta_0(\tau,\xi) = \frac{[\chi]_0(t,S)}{\Theta}, \quad \psi(\xi) = \frac{\Psi(S)}{\Theta}, \quad n(\tau) = \frac{\mathrm{N(t)}}{S_d\Theta}, \quad \gamma(\tau) = \lambda(t)\frac{T_d}{S_d\Theta}, \quad \gamma_{in}(\tau) = \lambda_{in}(t)\frac{T_d}{S_d\Theta}, \tag{7}$$

$$g(\tau) = a(t)\frac{T_d}{S_d}, \ \Theta = \max\left\{\Psi(S), \frac{\lambda(t)}{a(t)}\right\}, \delta(\xi) = S_d\delta(S), \ H(S_d\xi) = H(S), \qquad a(t) \neq 0. \tag{8}$$

The value of the specific load on the conveyor belt should not exceed the maximum permissible value

$$\frac{\lambda(t)}{a(t)} = [\chi]_0(t,0) \leq [\chi]_{0\max}.$$

With the dimensionless value $n(\tau) = 1.0$ and $\Theta = [\chi]_{0\max}$ the accumulating bunker contains the amount of material $\mathrm{N(t)} = S_d\Theta$, that will fill the conveyor line along the entire length with the maximum allowable rock distribution density $[\chi]_0(t,S) = [\chi]_{0\max}$. Taking into account the notation (6)–(8), the balance equation for the flow parameters of the conveyor line is written in the dimensionless form [21]:

$$\frac{\partial\theta_0(\tau,\xi)}{\partial\tau} + g(\tau)\frac{\partial\theta_0(\tau,\xi)}{\partial\xi} = \delta(\xi)\gamma(\tau), \qquad \theta_0(0,\xi) = \mathrm{H}(\xi)\psi(\xi), \tag{9}$$

$$\frac{dn(t)}{dt} = \gamma_{in}(\tau) - \gamma(\tau), \qquad n(0) = n_0, \qquad 0 \leq n(\tau) \leq n_{\max}, \qquad 0 \leq \gamma(\tau) \leq \gamma_{\max}. \tag{10}$$

The solution of the system of equations (9) has the form:

$$\theta_0(\tau,\xi) = \left[H(\xi) - H\left(-G(\tau_\xi)\right)\right]\frac{\gamma(\tau_\xi)}{g(\tau_\xi)} + H\left(-G(\tau_\xi)\right)\psi\left(-G(\tau_\xi)\right), \tag{11}$$

$$G(\tau) = \int g(\tau)d\tau, \qquad \tau_\xi = G^{-1}\left(G(\tau) - \xi\right). \tag{12}$$

For a constant speed of the conveyor belt $g(\tau) = g_0 = const$ expressions (12) can be represented as

$$G(\tau) = g_0\tau, \qquad \tau_\xi = \frac{G(\tau) - \xi}{g_0} = \frac{g_0\tau - \xi}{g_0} = \tau - \frac{\xi}{g_0}, \tag{13}$$

whence the expression for the distribution material density along the transportation route at an arbitrary point in time $\theta_0(\tau,\xi)$ along the transportation route at an arbitrary point in time τ

$$\theta_0(\tau,\xi) = \left[H(\xi) - H\left(\xi - g_0\tau\right)\right]\frac{\gamma\left(\tau - \frac{\xi}{g_0}\right)}{g_0} + H\left(\xi - g_0\tau\right)\psi\left(\xi - g_0\tau\right). \tag{14}$$

For the transportation process by $\tau > \xi/g_0$

$$\theta_0(\tau,\xi) = \frac{\gamma\left(\tau - \frac{\xi}{g_0}\right)}{g_0}, \qquad \theta_1(\tau,\xi) = \gamma\left(\tau - \frac{\xi}{g_0}\right), \qquad \tau > \frac{\xi}{g_0}. \tag{15}$$

In order to determine the value of linear density $\theta_0(\tau,\xi)$ or material flow $\theta_1(\tau,\xi)$ at an arbitrary point ξ at a point in time τ, it is required to know the value of the input material flow to the conveyor line at a point in time

$\tau_\xi = \tau - \xi / g_0$. The material flow at the output from the conveyor line $\theta_1(\tau,1)$ at a constant belt speed is determined by the product g_0 и $\theta_0(\tau,1)$

$$\theta_1(\tau,1) = g_0\theta_0(\tau,1) = g_0\frac{\gamma(\tau - 1/g_0)}{g_0} = \gamma(\tau - 1/g_0). \tag{16}$$

In general, when the belt speed and the input material flow are variable, the expression for the material flow at the exit of the conveyor section is

$$\theta_1(\tau,1) = g(\tau)\left[H(1) - H(-G(\tau_1))\right]\frac{\gamma(\tau_1)}{g(\tau_1)} + H(-G(\tau_{1\xi}))\psi(-G(\tau_1)), \qquad \tau_1 = G^{-1}(G(\tau) - 1). \tag{17}$$

Expressions (16), (17) determine the value of the output material flow from the conveyor section, which can be used to design a control system for the flow parameters of a conveyor-type transport system.

Results

1. The problem of optimal control of the conveyor belt speed

Let us formulate the problem of constructing an optimal program for controlling the speed of a conveyor belt for a steady state operation $G(\tau) - 1 > 0$ of a conveyor line: determine the material output $\theta_1(\tau,1)$ from the conveyor line during a period of time $\tau = [0, \tau_k]$ with step control of the speed of the conveyor belt $u(\tau) = (u_1, u_2)$, $0 < u_1 < u_2 < \infty$, $u_1 = const$, $u_2 = const$, which leads to a minimum functionality

$$\int_0^{\tau_k} |\theta_1(\tau,1) - \vartheta(\tau)| d\tau \to \min \tag{18}$$

with differential connections (9)

$$\frac{\partial\theta_0(\tau,\xi)}{\partial\tau} + u(\tau)\frac{\partial\theta_0(\tau,\xi)}{\partial\xi} = \delta(\xi)\gamma(\tau), \qquad g(\tau) = u(\tau) \tag{19}$$

and restrictions

$$\theta_0(\tau,\xi) \geq 0, \tag{20}$$

and initial conditions

$$\theta_0(0,\xi) = \mathrm{H}(\xi)\psi(\xi). \tag{21}$$

Let us assume that prior to the introduction of control, the conveyor operated in a steady state at a constant speed $g(\tau)\big|_{\tau<0} = u_1$. The Pontryagin function and the adjoint system have the form:

$$H = -\left|\frac{\gamma(\tau - \Delta\tau_1)}{u(\tau - \Delta\tau_1)}u(\tau) - \vartheta(\tau)\right| + \psi_1 u(\tau) \to \max, \qquad \int_{\tau-\Delta\tau_1}^{\tau} u(\tau)d\tau = 1, \tag{22}$$

$$\frac{d\psi_1}{d\tau} = \frac{\partial H}{\partial\xi} = 0. \tag{23}$$

Since the right end of the phase trajectory is free, then $\psi_1(\tau_k) = 0$ and, therefore, $\psi_1(\tau) \equiv 0$, which allows us to write the Pontryagin function in the following form

$$H = -\left|\frac{\gamma(\tau - \Delta\tau_1)}{u(\tau - \Delta\tau_1)}u(\tau) - \vartheta(\tau)\right| \to \max. \tag{24}$$

Let us construct the optimal control of the speed of the conveyor belt for the case of supplying material to the input of the conveyor with a constant intensity $\gamma(\tau)=1$ with the existing demand, which is determined by a periodic function $\vartheta(\tau)=1+\sin(\omega\tau)$ [22]. Let's believe that the conveyor line is capable of operating in one of the speed modes: with the speed of the belt u_1 or u_2, $u(\tau)=(u_1,u_2)$. The calculation results are shown in Fig. 2 for different values of the stepwise control of the belt speed u_1 or u_2: $u(\tau)=(0.5,2.0)$, Fig.2a; $u(\tau)=(0.8,2.0)$, Fig.2b; $u(\tau)=(1.0,2.0)$, Fig.2c; $u(\tau)=(1.2,2.0)$, Fig.2d; $u(\tau)=(1.8,2.0)$, Fig.2e. The maximization of function (22) determines such a control $u(\tau)$, which the output of products on the conveyor line is ensured with a minimum deviation from the existing demand $\vartheta(\tau)$. Figure 2 shows the calculation of the optimal control $u(\tau)$ of the speed of the conveyor belt, depending on the demand $\vartheta(\tau)$. Graphs (a–e) on the left side in Fig. 2 represent the dependence of optimal control $u(\tau)$ on the value of demand $\vartheta(\tau)$. On the right, for each variant of the dependence $u(\tau)$, $\vartheta(\tau)$ a graph of the family of characteristics with breaks at the time of switching the speed mode is presented. Each control mode (a–e), shown in Fig. 2, corresponds to the time dependence of the output flow $\theta_1(\tau,1)$ from the conveyor against the background of existing demand $\vartheta(\tau)$. For the step control speed option $u(\tau)=(0.5,2.0)$, Fig. 2a, there are no conveyor belt speed switching modes.

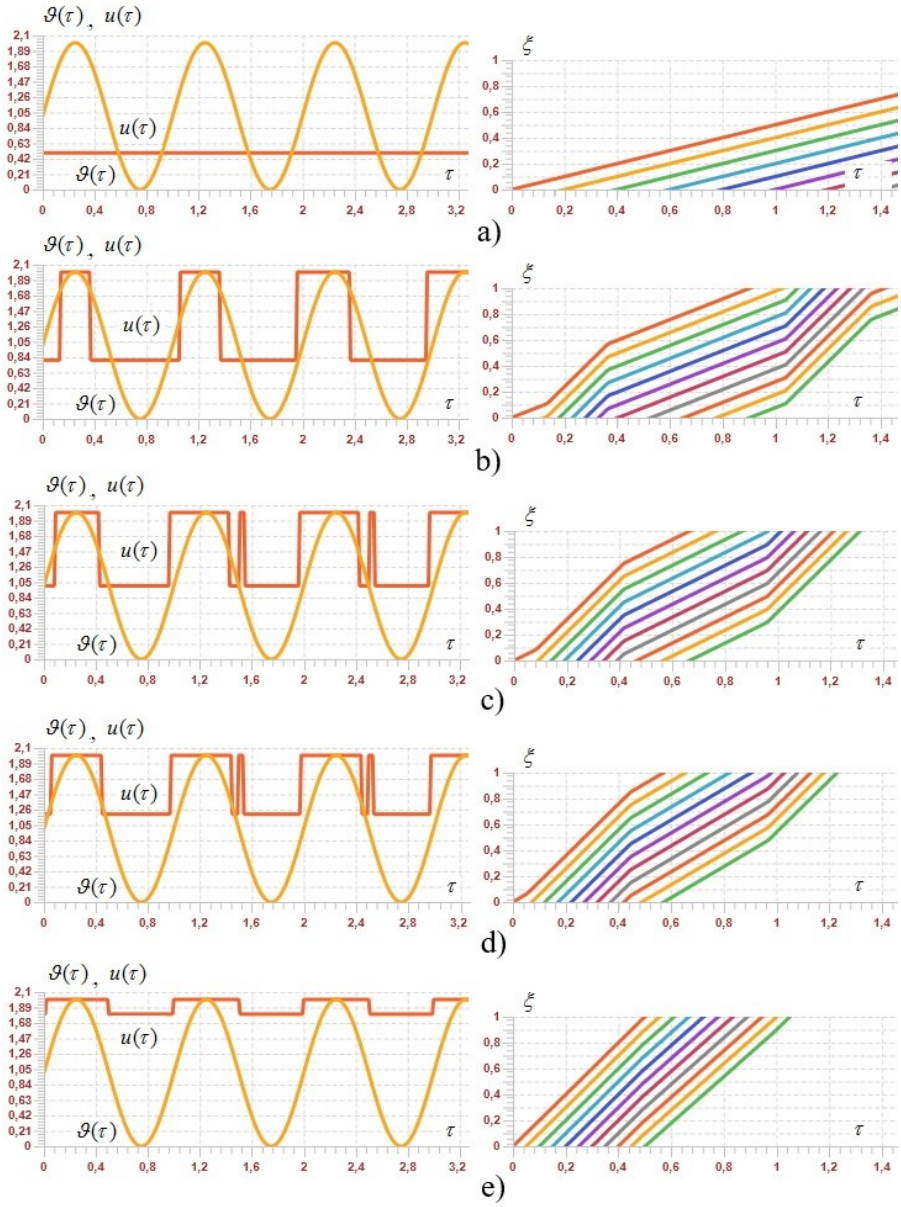


Fig.2. Optimal control and a family of characteristics for operating modes

The output material flow from the conveyor line $\theta_1(\tau,1)$ is constant and does not depend on the existing demand. This behavior is due, to some extent, to the fact that the initial speed of the conveyor line without control is

taken as the smaller of the two, $g(\tau)\big|_{\tau<0} = u_1$, and a significant spread between the step value of the control speeds, which is characterized by a coefficient (Fig. 2a) $k_u = u_2 / u_1 = 4.0$. When decreasing k_u by increasing the value u_1 the conveyor starts to operate in a two-speed mode (Fig. 2b), and the family of characteristics at these switching moments has kinks At the same time, the output flow from the section $\theta_1(\tau,l)$ is determined by the behavior of demand $\vartheta(\tau)$ and takes one of three values over time {0.4; 1.0; 2.5}, which is determined by the relation $\frac{u(\tau)}{u(\tau-\Delta\tau_1)} = \left\{\frac{2.0}{0.8}, \frac{0.8}{0.8}, \frac{0.8}{2.0}, \frac{2.0}{2.0}\right\}$.

The unevenness of material density $\theta_0(\tau,\xi)$ is set by the speed mode of the conveyor line and takes one of two values $\theta_0(\tau,\xi) = \frac{\gamma(\tau_\xi)}{g(\tau_\xi)} = \left\{\frac{1.0}{0.8}, \frac{1.0}{2.0}\right\}$. The duration of the operation of the section in the control mode $u(\tau) = u_2$ increases with each switch and reaches the steady state (Fig. 2b). With a subsequent decrease, short-term peak switchings are added (Fig. 2c, Fig. 2d), which disappear with a further decrease k_u (Fig. 2e).

This behavior of the control function is explained by the fact that the speed control $u(\tau)$ of the section depends on the accepted control $u(\tau-\Delta\tau_1)$, where is $\Delta\tau_1$ a time-dependent value estimated by relation (22).

2. Synthesis of an algorithm for controlling the speed belt based on TOU-tariffs

The problem of constructing an optimal program for controlling the belt speed is formulated as follows: determine the modes of switching the belt speed during a period of time $\tau = [0, \tau_{24}]$ with the value of the price coefficients of the cost of electricity $z(\tau)$ (Fig. 3) with stepwise control of the belt speed $g(\tau) = u(\tau) = (u_1, u_2)$, $0 < u_1 < u_2 < \infty$, $u_1 = const$, $u_2 = const$, which leads to a minimum of the functional:

$$\int_0^{\tau_{24}} z(\tau)u(\tau)m(\tau)d\tau \to \min, \tag{25}$$

with differential connections

$$\frac{dm(\tau)}{d\tau} = \gamma_1(\tau) - \theta_1(1,\tau) = \gamma_1(\tau) - \gamma_1(\tau-\Delta\tau_1)\frac{u(\tau)}{u(\tau-\Delta\tau_1)}, \tag{26}$$

$$m(0) = 2(\theta_{0R} + \theta_{0C}) + \int_0^1 \psi(\xi)d\xi, \tag{27}$$

and a limit on the total amount of energy consumed per day

$$\int_0^{\tau_{24}} u(\tau)m(\tau)d\tau = b = const. \tag{28}$$

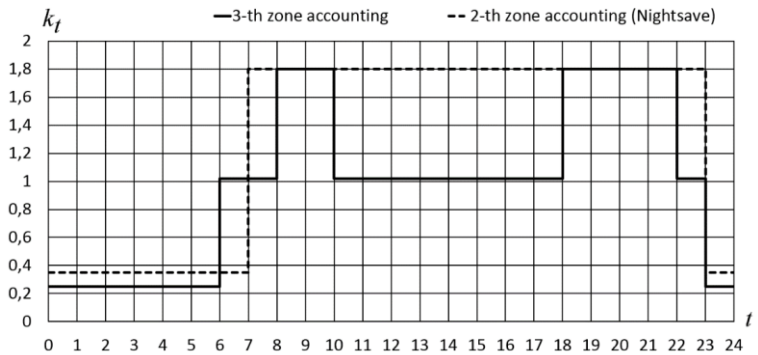


Fig.3. Tariff coefficients Ukraine–TOU periods (2020)

Equation (26) is written taking into account the dependence on the output material flow. The choice of a stepwise control mode is due to its prevalence in the control of transport systems [23, 24]. Equation (28) can be replaced by the differential equation

$$\frac{dx_b}{d\tau} = u(\tau)m(\tau), \qquad x_b(0) = 0, \qquad x_b(\tau_{24}) = b. \tag{28}$$

The Hamilton function and the adjoint system of equations for the problem under consideration has the form:

$$H = (\psi_b - z(\tau))u(\tau)m(\tau) + \psi_m\left(\gamma_1(\tau) - \gamma_1(\tau - \Delta\tau_1)\frac{u(\tau)}{u(\tau - \Delta\tau_1)}\right), \tag{29}$$

$$\frac{d\psi_b}{d\tau} = -\frac{\partial H}{\partial x_b} = 0, \tag{30}$$

$$\frac{d\psi_m}{d\tau} = (z(\tau) - \psi_b)u(\tau) \qquad \psi_m(\tau_{24}) = 0. \tag{31}$$

Equation (31) implies $\psi_b = C_b = const$. For the two-stage control mode $u(\tau) = (u_1, u_2)$ $0 < u_1 < u_2 < \infty$ the optimal belt speed corresponds to the maximum value of the Hamilton function (29). Switching points of control modes are determined by solving equations (26)-(31). To carry out quantitative calculations, let's take the value of the time parameter and the value of the sample $T_d = 1$ (*hour*), $S_d = 20.5$ (km). The choice of the characteristic time value T_d makes it possible to conveniently display the change in parameters during the day $\tau \in [0;24]$, and the choice of the characteristic length value corresponds to the consideration of extended transport conveyor (Sasol – Shondoni Overland (20.5 km single flight overland conveyor with multiple horizontal curves). Then the control modes $u(\tau) = (u_1, u_2)$, at speeds $a_1(\tau) = (1.0, 5.0)$ m/sec, will correspond to the dimensionless values of the belt speed $u(\tau) = (0.176,\ 0.878)$.

Let's consider the construction of a schedule for switching belt speed modes for tariff coefficients Ukraine–TOU periods (Fig. 3), when the intensity of material receipt is a constant value $\gamma_1(\tau) = 0.15$ with daily energy consumption $b = 6.5$. The selected value of the intensity of the incoming flow allows the transport system to operate in a two-speed mode $u(\tau) = (0.176,\ 0.878)$ (19). Increasing the belt speed $g(\tau)$ leads to an increase in the power consumption of the transport system $n_e(\tau)$. On the other hand, an increase in belt speed leads to a decrease in the linear density of the material $\theta_0(\tau,0)$, entering the section input and, possibly, to a decrease in the mass $m(\tau)$. The next feature is that the conveyor belt is an accumulator of the material entering the section input. The limitation does not allow both the excess of the specific gravity and the overflow of the accumulator. The presence of upper and lower limits for the amount of material in the accumulator with a sufficiently long period of time of the transport system operation determines the ratio between the average intensity of the incoming flow and the average power consumption of the transport system.

An increase in the value of the intensity of the incoming material over a sufficiently long period of time of the operation of the transport system leads to an increase in the power consumption of the transport system.Эти особенности значительно усложняют синтез алгоритмов оптимального управления потоковыми параметрами транспортной системы. In this regard, to simplify the qualitative analysis, a constant value for the intensity of the incoming flow $\gamma_1(\tau)$ is taken. Belt speed control modes $u(\tau) = (u_1, u_2)$ for tariff coefficient values $k_\tau(\tau) = (0.35,\ 1.8)$ are shown in Fig. 4.

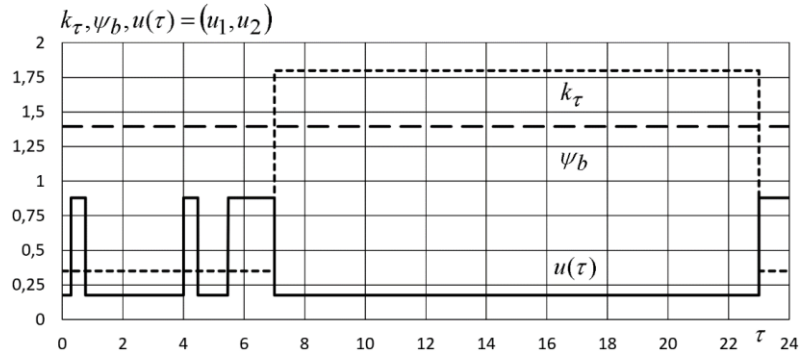


Fig 4. Belt speed control modes for tariff coefficient values $k_\tau(\tau) = (0.35,\ 1.8)$ **and** $\psi(\xi) = 0.5 + 0.5\sin(2\pi\xi)$

The value of the tariff coefficient $k_\tau = 1.8$ corresponds to the minimum belt speed u_1. There are several speed-switching points for the tariff coefficien $k_\tau = 0.35$. Switching on the interval $\tau \in [0.0;1.0]$ is related to the type of function (21), which determines the initial distribution of material along the route. Switching for the interval $\tau \in [4.0;5.0]$ repeats the switching mode for the interval $\tau \in [0.0;1.0]$ with a time shift equal to the amount of transport delay.

For the moments of time when the value of the transport delay has high values, the mass of the belt with the transported load also has high values (Fig. 4). An increase in the transport delay leads to an increase in the mass of material on the belt.

Conclusions

The article discusses a technique using the PiKh–model of a conveyor section for the synthesis of algorithms for optimal control of the belt speed for a distributed transport system of a conveyor type. A method for constructing an algorithm based on the Pontryagin maximum principle using an analytical PiKh–model for a conveyor section is proposed. The use of the Pontryagin maximum principle together with the analytical PiKh–model allows us to provide acceptable accuracy for calculating the switching points of speed modes. To carry out numerical experiments, software was used that makes it possible to synthesize algorithms for optimal discrete control of the belt speed for practical and theoretical purposes. The presented technique for synthesizing optimal control algorithms makes it possible to take into account the initial distribution of material along the transportation route when choosing control modes.

The main advantage of the presented technique is that the synthesis of optimal control algorithms takes into account the variable transport delay for the conveyor section.

The prospect for further research is the development of a methodology for synthesizing algorithms for optimal continuous control of the speed of a conveyor belt in the design of control systems for multi-section transport conveyors.

References

1. Pihnastyi O. Statistical theory of control systems of the flow production. Lap Lambert Academic Publishing, 2018. 436 p.
2. Fedorko G. Implementation of Industry conveyor transport. MATEC Web of Conferences 263, 01001. 2019. https://doi.org/10.1051/matecconf/201926301001
3. Siemens. Innovative solutions for the mining industry. 2018. www.siemens.com/mining.
4. He D., Pang Y., Lodewijks G., Liu X. Healthy speed control of belt conveyors on conveying bulk materials. *Powder Technology*, 2018. No 327, P. 408–419. http://dx.doi.org/10.1016/j.powtec.2018.01.002
5. Zhang S., Xia X. Modeling and energy efficiency optimization of belt conveyors. *Applied energy*, 2011. No 88(9), P. 3061–3071, 2011. http://nbuv.gov.ua/UJRN/Npdntu_gir_2013_2_15
6. Bajda M., Błażej R., Jurdziak L. Analysis of changes in the length of belt sections and the number of splices in the belt loops on conveyors in an underground mine. *Engineering Failure Analysis*, 2019. No 101, P. 439–446. https://doi.org/10.1016/j.engfailanal.2019.04.003
7. Pihnastyi O., Khodusov V. Neural model of conveyor type transport system. Proceedings of The Third International Workshop on Computer Modeling and Intelligent Systems, 2020. http://ceur-ws.org/Vol-2608/paper60.pdf
8. Razumnyj Ju., Ruhlov A., Kozar A. Improving the energy efficiency of conveyor transport of coal mines. *Mining Electromechanics and Automation*, 2006. No 76, P. 24–28.
9. Mathaba T, Xia X. A parametric energy model for energy management of long belt conveyors. *Energies*, 2015. No 8(12), P. 13590–13608. https://doi.org/10.3390/en81212375
10. DIN 22101:2002-08. Continous conveyors. Belt conveyors for loose bulk materials. Basics for calculation and dimensioning. [Normenausschuss Bergbau (FABERG), Deutsches Institut für Normung e.v. Normenausschuss Maschinenbau (NAM)], 2002.
11. Kiriia R., Shyrin L. Reducing the energy consumption of the conveyor transport system of mining enterprises. *Essays of Mining Science and Practice*: International Conference. 2019. https://doi.org/10.1051/e3sconf/201910900036
12. Marais J, Mathews E, Pelzer R. Analysing DSM opportunities on mine conveyor systems. *Industrial and commercial use of energy* conference. 28–30 May 2008, Cape Town, South Africa. 2008.
13. Pingyuan Xi, Yandong Song. Application Research on BP Neural Network PID Control of the Belt Conveyor. *JDIM*, 2011. No 9(6), P. 266–270
14. Razumnyj Ju., Ruhlov A., Kozar A. Increased energy efficiency conveyor transport of coal mines. *Girnicha elektromehanika ta avtomatika*. 2006. No 76. P. 24–28. https://vde.nmu.org.ua/ua/science/ntz/archive/76/5.pdf
15. Wolstenholm E. Designing and assessing the benefits of control policies for conveyor belt systems in underground mines. *Dynamica*, 1980. No 6(2), P. 25–35. https://www.systemdynamics.org/assets/dynamica/62/6.pdf
16. Andrejiova M., Marasova D. Using the classical linear regression model in analysis of the dependences of conveyor belt life. *Acta Montanistica Slovaca*, 2013. No 18(2), P. 77–84. https://actamont.tuke.sk/pdf/2013/n2/2andrejiova.pdf
17. Xinglei L., Hongbin Yu. The Design and Application of Control System Based on the BP Neural Network. *Mechanical Engineering and Intelligent Systems*: III International Conference, 2015. P. 789–793. https://doi.org/10.2991/icmeis-15.2015.148
18. Pihnastyi O.M. Analytical methods for designing technological trajectories of the object of labour in a phase space of states. *Scientific bulletin of National Mining University*. 2017. No 4. P. 104–111. http://nvngu.in.ua/index.php/en/component/jdownloads/viewdownload/69/8679
19. Khodusov V.D. The use of physical kinetics methods for studying oscillations of the parameters of a production line. East *European journal of physics*. 2014. No 1(4). P 88-95. http://nbuv.gov.ua/UJRN/eejph_2014_1_4_12
20. ConveyorBeltGuide. Example of use. 2018. –Available at http://conveyorbeltguide.com/examples-of-use.html. (accessed 2 December 2022)
21. Pihnastyi O.M. Distinctive numbers of production systems functioning description. *Problems of Atomic science and technology*. 2007. No3. P. 322–325. http://dspace.nbuv.gov.ua/handle/123456789/111018

22. Nway Hlaing, Ei Htwe, Maw Htay Dynamic Analysis of the Belt Conveyor System for Detergent. *International Journal of Scientific and Research Publications*. 2018. No 8(10). P. 633–639. https://www.ijsrp.org/research-paper-0918.php?rp=P827889

24. Lauhoff, H. Speed Control on Belt Conveyors–Does it Really Save Energy? *Bulk Solids Handling*. 2005. No 25(6). P. 368–377. http://synergy-eng.com/pdf/BSH-2005_Beltspeed_Lauhoff.pdf

25. Halepoto, M. Uqaili, Design and Implementation of Intelligent Energy Efficient Conveyor System Model Based on Variable Speed Drive Control and Physical Modeling. *International Journal of Control and Automation*. 2016. No 9(6). P. 379–388. http://article.nadiapub.com/IJCA/vol9_no6/36.pdf

Oleh Pihnastyi **Олег Пігнастий**	Doctor of Engineering Science, Professor, Department of distributed information systems and cloud technologies, National Technical University "Kharkiv Polytechnic Institute" e-mail: pihnastyi@gmail.com https://orcid.org/0000-0002-5424-9843	доктор технічних наук, проф., каф. розподілених інформаційних систем та хмарних технологій, Національний технічний університет «Харківський політехнічний інститут», Харків, Україна
Maksym Sobol **Максім Соболь**	PhD, Department of Computer Science and Intellectual Property, National Technical University "Kharkiv Polytechnic Institute" e-mail: maksym.sobol@khpi.edu.ua https://orcid.org/0000-0002-7853-4390	кандидат технічних наук, старший викладач, каф. інформатики та інтелектуальної власності, Національний технічний університет «Харківський політехнічний інститут», Харків, Україна

UDK 336:338.4+519.7
https://doi.org/10.31891/csit-2022-4-11

LARYSA SAVCHUK, LILIIA BANDORINA, KATERYNA UDACHINA
Ukrainian State University of Science and Technologies
IRYNA SOKYRYNSKA
Zahidnodonbaskyy Institute of IAPM
ROMAN SAVCHUK
Ukrainian State University of Science and Technologies

DEVELOPMENT OF SOFTWARE MODULE FOR DETERMINING RATIONAL CAPITAL STRUCTURE OF THE ENTERPRISE

The relevance of research on the use of information technologies to solve applied problems, in particular, financial policy management at the micro level has been revealed in this article. The own definition of the concept of financial policy at the enterprise level has been proposed as a triad of the main content principles, namely: asset management, cash flow management, liability management. It has been established that research in the field of financial policy is mostly theoretical, therefore the development of software module based on the model of solving a practical problem presented in the article is relevant for Ukrainian enterprises.

After structurization and determination the list of financial policy tasks, an original improvement of one of the important components in asset management, namely the task of modeling the amount of receivables, has been proposed. In order to obtain the maximum economic effect, it has been proposed to suggest to calculate the optimal level of accounts receivable. Using of the method of approximation of the non-linear dependence of receivables on the credit period, which was initiated by O. I. Luchkov, has been substantiated. The principles of the methodics have been defined, its modification and clarification have been proposed in the direction of the development of methods for calculating the numerical values of all components of the model in the article. This method allows the obtaining the optimal term of the credit period in days and the optimal amount of receivables in monetary terms.

Model experiment has been executed to test the model's operation, during which data from the financial statements of one of the operating enterprises have been used. The obtained results testify to the effectiveness and correctness of the proposed methodics are served as an information basis for management decisions regarding the assessment of the effectiveness of measures to manage the company's financial policy.

The software module has been developed that, based on the results of certain calculations, allows to make quick and well-founded decisions about balancing the structure of assets and liabilities in order to restore financial stability and increase profits.

Keywords: software module, information technology, financial policy, triad, asset management, receivables, approximation of nonlinear dependence, model experiment

ЛАРИСА САВЧУК, ЛІЛІЯ БАНДОРІНА, КАТЕРИНА УДАЧИНА
Український державний університет науки і технологій
ІРИНА СОКИРИНСЬКА
Західнодонбаський інститут ПрАТ « ВНЗ Міжрегіональна академія управління персоналом»
РОМАН САВЧУК
Український державний університет науки і технологій

РОЗРОБКА ПРОГРАМНОГО МОДУЛЯ ДЛЯ ВИЗНАЧЕННЯ РАЦІОНАЛЬНОЇ СТРУКТУРИ КАПІТАЛУ ПІДПРИЄМСТВА

У статті розкрито актуальність дослідження питань використання інформаційних технологій для вирішення прикладних задач, зокрема, управління фінансовою політикою на мікрорівні. Надано власне визначення змісту поняття фінансової політики на рівні підприємства як тріади основних змістовних засад, а саме: управління активами, управління грошовими потоками, управління пасивами. Виявлено, що дослідження у сфері фінансової політики переважно теоретичні, тому розробка програмного модуля на основі моделі вирішення практичної задачі, що надається у статті, є актуальною для підприємств України.

Після структурування і визначення переліку задач фінансової політики запропоновано оригінальне удосконалення однієї з важливих складових в управлінні активами, а саме задачі моделювання розміру дебіторської заборгованості. З метою отримання максимального економічного ефекту запропоновано розрахувати оптимальний рівень дебіторської заборгованості.

Обґрунтовано використання методики апроксимації нелінійної залежності дебіторської заборгованості від кредитного періоду, що була започаткована О. І. Лучковим. У статті чітко визначені всі засади методики, запропонована її модифікація і уточнення у напрямку розробки методів розрахунку числових значень всіх складових моделі. Даний метод дозволяє отримати оптимальний термін кредитного періоду в днях та оптимальний обсяг дебіторської заборгованості у грошовому еквіваленті.

Для перевірки роботи моделі здійснено модельний експеримент, при проведенні якого були використані дані фінансової звітності одного із діючих підприємств. Отримані результати свідчать про дієздатність і коректність запропонованої методики і виступають інформаційним підґрунтям для управлінських рішень стосовно оцінки ефективності заходів по управлінню фінансовою політикою підприємства.

Розроблено програмний модуль, який за результатами певних розрахунків дозволяє приймати оперативні та обґрунтовані рішення стосовно балансування структури активів й пасивів для відновлення фінансової стійкості та нарощування прибутку.

Ключові слова: програмний модуль, інформаційні технології, фінансова політика, тріада, управління активами, дебіторська заборгованість, апроксимація нелінійної залежності, модельний експеримент.

Introduction

The use of information technologies to solve applied business problems is an actual problem for many entrepreneurs. The implementation of the software contributes to the reduction of time for data processing, allows quickly and reasonably make appropriate decisions, which affects the overall effectiveness of the work of organizations.

An important role in the activity of the enterprise is the economic subsystem, namely: the issue of effective management of the capital structure of the enterprise, which is a guarantee of the stability of its development, ensuring its solvency, strengthening the financial condition, increasing the market value and the wellbeing of the owners. The capital structure determines the possibilities of the enterprise to form the composition of its assets, provided that a safe level of financial stability is maintained.

Analysis of the actual state of sources of finance and, accordingly, the capital structure of enterprises shows that their financial resources are not used efficiently enough. This is a consequence of the irrational distribution of the company's funds and the source of their financing. The formation of a rational capital structure should begin with the determination of factors that affect the current financial state, using this selected algorithm for calculating indicators. The detection of the magnitude of the further change of the factor and the possibility of its change should determine the further financial strategy of the enterprise, and the proposed recommendations for choosing ways to improve the financial state and stabilize its activity in the future.

The structure of investment capital plays a role in the formation of the market value of the enterprise. The relationship is determined by the weighted average cost of capital. Therefore, it is expedient to explore the concept of capital structure together with the concept of market value of enterprise, and to improve the model for determining the rational capital structure.

Usually a company can choose any capital structure it wants. But the main management purpose of the enterprise is the choice of such a structure to maintain stable income and dividends at the lowest capital value.

That is, the optimal capital structure should minimize the average capital value and at the same time maintain the credit reputation of the enterprise at a level that would allow attracting new types of capital on terms acceptable to this enterprise.

Management of the financial component of capital is based on a process that can be formalized by defining a certain sequence of actions during managing the capital structure..

The purpose of capital structure management is to find the optimal ratio between equity and debt capital, various short-term and long-term sources of its formation (the share in the total amount of equity liabilities, long-term liabilities, short-term liabilities in the form of bank loans, credit lines, accounts payable, etc.) [1].

The optimal capital structure is the result of a compromise between the achievement of the maximum possible the tax savings due to the attraction of borrowed funds and additional costs associated with an increase in the probability of financial difficulties with an increase in the share of loan capital [2, 3].

Related Works

Capital is one of the fundamental economic categories, the meaning of which scientific thought has been investigating for many centuries [2, 3]. 4,5

Among the theoretical foundations of capital management formation, one of the basic ones is the theory of its structure. This is associated with the fact that the theoretical concept of the capital structure forms the basis for choosing a number of strategic directions of the financial development of the enterprise, which ensure the growth of its market value [6]. Therefore, this theory is a wide enough field of practical application.

The definition of capital structure is disputable and so needs clear definition.

With condition of widespread practical use of the definition of capital structure, many economists proposed to expand the composition of loan capital by adding various types of short-term bank credit to it. They associated the possibility of approach to the concept of "capital structure" with the increased role of bank credit in financing the economic activity of enterprises and the expanded practice of restructuring short-term to long-term types of it [7].

Many specialists think that the definition of "capital structure" it is necessary to consider all types of both equity and loan capital of the enterprise in current conditions.

At the same time, the composition of equity capital should include not only the initially invested amount (shareholder capital which participates in the formation of the company's statutory fund), but also the share in the form of certain reserves and funds created in the activities of the enterprise, and retained earnings . Based on this, loan capital must be considered according to the forms of use, taking into account financial leasing, commodity credit, accounts payable, etc.

This interpretation of the concept of "capital structure" allows to significantly expand the area of practical application of this theoretical concept in the financial activity of an enterprise, because [8]: it effectively allows to investigate the specifics and develop appropriate recommendations for enterprises with limited financing opportunities on the long-term capital market; makes it possible to connect the capital structure with the use of assets financed by its various components.

Taking into account the considered provisions, the concept of "capital structure" has the following definition: "capital structure is the ratio of all forms of own and borrowed funds used by the enterprise for the purpose of financing assets" [2].

The formation of the financial structure of the capital as a way of financing the activity of the enterprise as a whole occupies an important place in the theory of financial management in the process of managing the capital of the enterprise, as it is complex factor that is interdependent with many indicators and characteristics of the economic activity of the enterprise.

Researchers have different views to issues related to capital structure. Different views on this problem led to the emergence of five stages in the development of the theory of capital structure, and, accordingly, five theoretical concepts [3, 9-11]:

- the concept of indifference of the capital structure;
- traditionalist concept of capital structure;
- compromise concept of the capital structure;
- the concept of "subordination of sources";
- the concept of conflicts of interests in the formation of the capital structure.

The first step towards the development of the theory of the capital structure was the hypothesis was proposed in 1958 by the American economists M. Miller and F. Modigliani, which was called the concept of indifference of the capital structure. According to this concept, the market value of the firm and the value of capital under certain conditions do not depend on its structure, which means that they cannot be optimized and the market value cannot be increased by changing the capital structure [12].

According to the founders of this theory, the cost of a corporation's capital is determined solely by its profit stream, which depends on investments. To justify this approach, the authors used a number of restrictions, some of which were later weaken.

Representatives of the traditional concept believe that, firstly, the cost of capital enterprise depends on its structure; secondly, there is an optimal capital structure that minimizes the value of the weighted average cost of capital and, therefore, maximizes the market value of the enterprise [13]. Based on this, an increase of specific weight of loan capital in all cases leads to decrease of weighted average cost of capital, and, accordingly, to an increase of the market value of the enterprise. The use of this concept in practice encourages the enterprise to increase of use of loan capital in economic activity, which can lead to negative consequences under certain conditions.

Having considered the advantages and disadvantages of these concepts of capital structure formation, we can conclude that it is expedient for the company to use loan capital until the tax benefits from the additional amount of credits and loans are equal to the costs associated with the growth of debt service and the increased probability of financial complications. This theory was formulated by S. Ross and was called the "static" theory of the capital structure [12].

Selection of previously unsolved parts of the overall problem

Considering the modern market of information systems, it is possible to conclude about the insufficiency of application software for determining the optimal capital structure of enterprise based on the model of the formation of the financial structure of capital depending on the property structure.

Purpose

The purpose of the article is the systematization of methodical approaches to the formation of the capital structure of the enterprise and the development of software module based on an advanced model of the formation of the financial structure of the capital depending on the property structure.

Results

The economic independence of enterprises in modern economic conditions is expressed by the size and structure of their capital. So there is quastion about at the expense of what property, at the expense of what sources (own or borrowed) of its acquisition and in what ratio should be formed the total capital of separate enterprise? Its solution requires only an individual approach [12-16].

Only optimal approach to the ratio of structural elements of capital for different enterprises or for one enterprise for the entire term of its operation cannot be determined. However, the process of choosing a rational capital structure of the enterprise should be carried out on the basis of the formation of the target capital structure.

So the dilemma of an enterprise's choice of one or another structure of its capital can be resolved only under the condition of choosing a certain criterion characteristic of optimality of the ratio of structural elements of capital. Different authors interpret the optimality of capital structuring differently according to certain criterion indicators. [17].

For example, such authors as Van Horn, Podderyogin, Ross, [8, 12, 18] define the optimal capital structure as a certain ratio of equity and borrowed capital, which leads to the maximization of the value of the company's shares.

That is, as a criterion for determining the optimality of the capital structure, these authors determine the maximum price of shares of enterprises.

The second group of authors determine value of the enterprise (firm) as a criterion for optimizing the ratio of own and borrowed capital . This group includes R. Clark, V.V. Kovaleva, O.O. Tereshchenko and others. Thus, according to these authors, the goal of optimizing the capital structure of the enterprise is the formation of an optimal ratio of various sources of financing the enterprise's activities, which maximizes the value of the enterprise itself [9, 19].

During investigation of the criterion for determining the optimality of the capital structure of the enterprise, some authors emphasize the selection of another criterion aspect, which maximizes the value of the enterprise. This is the ratio of profitability of own capital and financial stability of the enterprise. This criterion was highlighted by such authors as Blank, Ivko, Kvasnytska, Kostyrko and others [2, 9, 16, 17]. Blank I.O. notes that the optimal capital structure is such a ratio of the use of own and borrowed funds that ensures the most effective proportionality between the coefficient of financial profitability and the coefficient of financial stability of the enterprise, that is, its market value is maximized [2, 213 - 214].

Thus, the problem was considered the most detailed by I.O. Blank. In particular, it was highlighted [2, 213 - 214]:

– optimization method of the capital structure based on the criterion of maximizing the level of predictable financial profitability (own capital profitability). This method is based on multivariate calculations of the level of financial profitability (own capital profitability with different capital structure). The latter is expressed through such an indicator as the effect of financial leverage;

– optimization method of the capital structure based on the criterion of minimizing its cost. The method is based on a previous assessment of own and borrowed capital under various conditions of their formation, maintenance and realization of multivariate calculations of the weighted average cost of capital and, thus, finding the most realistic market value of the enterprise;

– optimization method of the capital structure based on the criterion of minimizing the level of financial risks. It is connected with the process of differentiated selection of funding sources for various components of the company's assets: non-current assets, current assets (with emphasizing of permanent and variable parts).

During choosing a rational capital structure of enterprise, it is necessary to achieve such a capital structure that will lead to the lowest value of the weighted average cost of capital at the same time as minimizing the risk of excessive dependence on short-term loan funds. The prposed model is based on the optimization model of the target capital structure of Raisa Kvasnytska, head of the Department of Finance and Banking of Khmelnytskyi National University [3, 11].

Therefore, the rational capital structure of the enterprise will be achieved while simultaneously minimizing the weighted average cost of capital and minimizing the risk of excessive dependence on current liabilities. The main measures, the purpose and the expected effect of the change in the capital structure are presented in table 1.

Table 1

Purpose, measures and effect of choosing a rational capital structure

Measures	Purpose	Effect
Increase of the share of loan capital	Using the effect of tax savings	Minimization the weighted-average cost of capital
Reducing of the share of short-term loan capital	Reducing the risk associated with mainly use of current liabilities.	Minimization the risk of excessive dependence on short-term loans

To reduce the weighted average cost of capital of the enterprise, it is necessary to increase the share of loan capital. The coefficient of financial leverage shows the ratio of the amount of loan capital to equity and is directly proportional to the financial risk of the enterprise. However, for the correct calculation of the choice of the capital structure, a large level of the financial leverage ratio determines a decrease of weighted average cost of capital and at the same time, as a rule, leads to deteriorate of the financial condition of the enterprise. That is, it is necessary to calculate the level of financial leverage not only in the structure of sources of financing of the enterprise's property (in the financial structure of capital), but also in the assets structure (in the property structure of the enterprise) [3, 11].

As financial stability is related to the possibility of financial support for repaying the company's debts, so determining ways to optimize the financial rational structure of capital, it is advisable with the study of the features of the financial structure of capital (which is the ratio of own and borrowed capital) to evaluate its property structure.

It is the rational structure of the company's property that is a positive factor for increasing the level of financial stability of the company. What is meant by the property structure of an enterprise's capital? Although the property ranking of the company's capital can be carried out in various aspects, in our opinion, the property structure of the capital should be considered in terms of all assets in circulation and property in non-monetary form.

Current assets are cash itself (both in national and foreign currency), short-term (current) financial investments and receivables. Non-monetary property form are production stocks, work-in-progress, finished products and goods.

Basis for calculating the financial security of covering liabilities in relation to the borrowed capital of the enterprise is achieved by separating current assets from non-monetary assets [3, 11].

So, in the financial aspect, we will consider the financial leverage of the property structure of the capital (the ratio of own and borrowed capital), in the property aspect it will be the financial leverage of the property structure of the capital (the ratio of circulate assets and in non-monetary assets).

It is necessary to determine the values of the group of indicators of the company's solvency for correct conclusions of the formation of both the property and financial structure of the capital. In our opinion, when choosing a rational capital structure of an enterprise based on the definition of the level of financial support for its obligations, the dominant indicator (exactly for our study) is the ratio of quick (critical) liquidity.

The indicator of quick liquidity, or "acid test", shows how all current financial obligations of the enterprise are secured by means of payment that it has on a certain date. Payment Means when calculating this indicator are cash, current financial investments and accounts receivable. It is worth noting that for our study these are circulate assets. The quick liquidity ratio is calculated as:

$$QLR = \frac{Cash + CFI + R}{CL}, \quad (1),$$

where QLR – quick liquidity ratio, CFI – Current financial investments, R – Receivables, CL – поточні зобов'язання.

It is on the basis of the calculation of this indicator we can evaluate the compliance of the share of the company's assets in circulation with certain regulatory values. The value of this indicator should be between 0.8 and 1.2 for the metallurgical industry.

Thus the following inequality is [20, 21]:

$$0.8 \le QLR \le 1.2, \quad (2)$$

It can be highlighted also the limit of the amount of assets in circulation, which is the most favorable for conducting economic activity based on the limit values of the critical liquidity ratio. After certain calculations formula is [4, 13]:

$$0.8 \le \frac{CA}{CL} \le 1.2 \quad (3)$$

or

$$CL * 0.8 \le CA \le 1.2, \quad (4)$$

де CA –активи в обігу.

The obtained inequality allows to evaluate the optimality of providing the enterprise by cash, current financial investments and receivables.

So, it is clear that for the criteria chosen by us for determining the rational capital structure of the enterprise, the level of financial leverage in the financial structure of the capital must prevail over the level of financial leverage of its property structure:

$$FLF > FLP, \quad (5)$$

where FLF – financial leverage if the financial structure;

FLP – financial leverage of property structure.

The calculation of financial leverage in the property structure of the enterprise's capital is carried out according to the following formula:

$$FLF = \frac{LC}{EC}, \quad (6)$$

де LC – loan capital;

EC –equity capital.

The level of financial leverage in the property structure of the enterprise's capital is calculated as [17, 20, 21]:

$$FLP = \frac{CA}{NA}, \tag{7}$$

where NA – non-monetary assets.

Financial leverage in both the financial and property capital structures based on the above formulas is calculated as:

$$\frac{LC}{EC} > \frac{CA}{NA}, \tag{8}$$

Thus, using this model, Ukrainian enterprises are able to form a rational capital structure for the combined effect of minimizing the weighted average cost of capital and reducing the risk of excessive dependence on short-term loan funds.

But, as was proven by previous calculations, the use of only this model for the formation of the optimal capital structure of the enterprise is insufficient, since the ratio between circulate assets and property in non-monetary form consists with a significant advantage in favor of the former - they exceed by 6.4 times assets in material form. It gives the enterprise certain additional opportunities to soften the financial structure of the capital in favor of loan sources. But it led to critical risky level for studied company: loan funds exceed the amount of equity capital by 13.7 times.

As to the property structure of the capital. It is determined by the features of the enterprise's economic activity. As to the financial structure, the sources of financing and the capital structure should be formed taking into account the property structure of assets.

Based on this, we believe that it is advisable to add clarifications to the model of R. Kvasnytska regarding the amount of own current assets, because a significant amount of current liabilities has led to lack of own working capital.

Since the main condition for ensuring an acceptable level of financial stability of enterprises in trade areas should be the presence of a positive amount of own capital and own working capital, we will analyze the trends of their changes.

Equity working capital (EWC) characterizes that part of equity capital and equivalent funds that is used to finance current assets. And although scientists discuss about the place of this indicator, in trade its role is important given the peculiarities of the field related to the volume, condition and dynamics of current assets.

In summary, we believe that R. Kvasnytska's model should be supplemented with the condition of having a positive value of own working capital. That is, for the studied company the conditions of formation of the optimal capital structure, taking into account the peculiarities of the property structure of its assets, it will be following conditions:

$$\begin{cases} FLF > FLP, \\ CL * 0.8 \le CA \le CL * 1.2 \\ OWC > 0, \end{cases} \tag{12}$$

where OWC – own working capital.

Thus, including additional restriction the risk of loss of financial stability will not be excessive.

Also compliance with the specified conditions is possible provided rational proportions are maintained between assets in circulation and short-term loan funds. The most active and influential components of both of these indicators are receivables and payables, respectively. At the same time, with a synchronous decreasing of the amount of equity due to losses, this control should be more stringent. After all, the growth of current liabilities against the background of a sharp drop of financial results leads to excessive risk of the capital structure.

Verification of the model was performed using a model experiment. Financial data of one of the operating enterprises were used.

Program module was developed based on proposed model, which startes to work with the main form. After entering the data, the rational capital structure is calculated. Figure 1 shows the external design of the main form with the results of the model.

After entering the input data in the table and pressing the "Calculation" button, calculated indicators of the company's capital structure assessment will be obtained, which will be placed in the appropriate fields.

The application is created in the Microsoft Visual Studio environment. This product and the tools included in it allow you to develop both console programs and programs with a graphical interface, including those with support for Windows Forms technology, which was used during the development of this application.

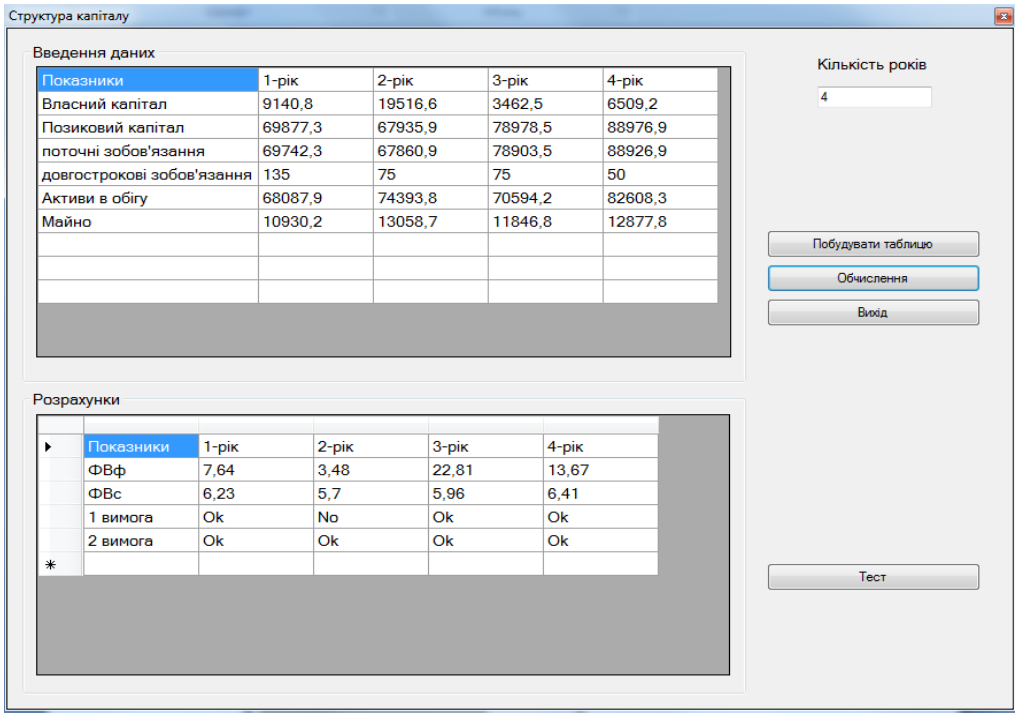


Fig. 1. The main modul form with results

The obtained results serve are information basis for management decisions regarding the optimal amount of receivables at the enterprise.

Conclusions

Software module was created in the work to increase the efficiency of the company's activities. This module will allow to make operational decisions regarding the management of the company's assets, namely, regarding the balancing of the structure of assets and liabilities to restore financial stability and increase profits.

The model of R. Kvasnytska was chosen as the basis of this system. The essence of the model is that all property is divided into circulate assets and assets in non-monetary form.

It has been assumed that the circulate assets are financed at the expense of loan sources, and property in non-monetary form are financed at the expense of own funds. For this purpose, the financial leverage of the financial structure and the financial leverage of the property structure are calculated, and additional restrictions are introduced regarding the ratio of circulate assets and loan funds

Approbation of the model on the data of the investigated enterprise proved that despite the fulfillment of all requirements, the capital structure remains excessively risky. So int the article proposes to improve the model by introducing an additional limitation is the need to have a positive amount of own working capital. So improved model with additional restrictions was proposed. Modeling on the basis of the proposed model made it possible to form such a system of management of funding sources, which minimizes financial risks even with the predominance of short-term loan funds.

Range of short-term loan funds has been formed for studied company based on proposed model, wthin which financial stability will be remain acceptable even with a small amount of equity capital.

Thus, the application of the created software module based on the proposed model in the process of work will make it possible to clearly and timely assess the level of the company's receivables.

References

1. Tatar M.S. Financial activity of business entities: education. manual. Kharkiv : KhAI, 2016. 99 p.
2. Blank Y.A. Financial management. K. : Эlha, Nyka_Tsentr, 2007. 521 p.
3. Kochkodan V.B. Formation and selection of measures to manage the capital structure of the enterprise. *Modeling of the regional economy: coll. of science works*. Ivano-Frankivsk : Play, 2007. No. 1(9). pp. 54-69.
4. Savchuk L., Kovalchuk K., Vyshniakova I., Sokyrynska I. Banking operations. K. : "Center for Educational Literature", 2013. 326 p.
5. Savchuk L., Udachyna K., Savchuk R. Modeling the optimal structure of the company's receivables. *Economic cybernetics: computer technologies in business*: coll. of science works based on the materials of the All-Ukrainian internet conference, Dnipro, March 1–2, 2022. Dnipro : UDUNT, 2022. pp. 164–169. URL: https://nmetau.edu.ua/file/zbirnik_2022.pdf
6. Kalinichenko 3.D. Using the leverage method in managing the profit and capital structure. *Economic analysis*. 2008. № 2 (18). pp. 212-214.
7. Tereshchenko O.O. Financial activity of economic entities. K. : KNEU, 2003. 554 p.
8. Podderyogin A.M. Financial management. Textbook / Curr. number author and science ed. Prof. A.M. Podderyogin K. : KNEU, 2005. 535 p.

9. L.A. Kostyrko L.A. The financial mechanism of sustainable development of enterprises: strategic guidelines, support systems, adaptation: monograph. Luhansk : "Knowledge" Publishing House, 2012. 474 p.
10. Kovalenko L.O. L.M. Remnova. Financial management. K. : Znannia, 2005. 485 p.
11. Krylenko V. I. Korneva N. O. Financial activity of economic entities: teaching method. manual. Kharkiv : Panov, 2016. 182 p.
12. Ross S., Westerfield R., Jaffe J., Jordan B. Corporate finance. McGraw Hill, 2018. 1040 p.
13. J. Collis, A. Holt, R. Hussey. An Introduction to Financial and Management Accounting. Palgrave Macmillan, 2012. 472 p.
14. Savchuk L., Bandorina L., Udachyna K., Savchuk R. A model for choosing a financial strategy for enterprise development. *Approaches to solving the problems of financing, security, economic and informational development of economic entities in the conditions of a pandemic: monograph* / edited by L.M. Savchuk, L.M. Bandorina. Dnipro. Porogy, 2021. pp. 265-285.
15. Savchuk L. M., Lozovska L. I., Savchuk R. V., Yarmolenko L. I. Modeling of problems of managing the financial policy of the enterprise. *Economic Bulletin of the Dnipro Polytechnic*. 2021. No. 1(73). pp. 170-178.
16. Ivko A.V. A comprehensive indicator of the optimal target structure of the enterprise's capital. *European vector of economic development*. 2011. No. 2 (11). pp. 105-110.
17. Kvasnytska P.C. Toolkit and methods of optimizing the target capital structure of the enterprise. *Economist*. 2005. No. 5. pp. 73-75.
18. Van Horne.Fundamentals of financial management. Pearson. 2008. 719 p.
19. Strategic financial management. Roger G. Clarke [et al.]. Homewood, Illinois : Irwin, 1988. 241 p.
20. Kvasnytska R.S. Ensuring the financial stability of the enterprise by the rational structure of its capital. *Bulletin of the Khmelnytskyi State University*. 2004. No. 1, Part 2, Volume 2. pp. 204-209.
21. Kvasnytska R.S. Causal factors of the dynamics of the target capital structure of enterprises as a potential for increasing the efficiency of their entrepreneurial activity. *Bulletin of the Lviv State Financial Academy: Collection of scientific articles*. 2005. No. 7. pp. 152-159.

Larysa Savchuk **Лариса Савчук**	Candidate of Economic Sciences, Professor, Dean of Faculty of Applied Computer Technologies, Ukrainian State University of Science and Technologies, Dnipro, Ukraine, e-mail: l.m.savchuk@ust.edu.ua http://orcid.org/0000-0003-2603-7218	к.е.н., професор, декан факультету прикладних компютерних технологій, Український державний університет науки і технологій, Дніпро, Україна
Liliia Bandorina **Лілія Бандоріна**	Candidate of Economic Sciences, Associate Professor, Head of the Department of Economic Informatics, Ukrainian State University of Science and Technologies, Dnipro, Ukraine, e-mail: l.m.bandorina@ust.edu.ua http://orcid.org/0000-0003-1383-3098	к.е.н., доцент, зав. каф. економічної інформатики, Український державний університет науки і технологій, Дніпро, Україна
Kateryna Udachyna **Катерина Удачина**	Candidate of Economic Sciences, Associate Professor of the Department of Economic Informatics, Ukrainian State University of Science and Technologies, Dnipro, Ukraine, e-mail: k.o.udachina@ust.edu.ua http://orcid.org/0000-0002-0043-6010	к.е.н., доцент, доцент кафедри економічної інформатики, Український державний університет науки і технологій, Дніпро, Україна
Iryna Sokyrynska **Ірина Сокиринська**	Candidate of Economic Sciences, Associate Professor of the Department of Economic and Management, Zahidnodonbaskyy Institute of IAPM, Pavlograd, Ukraine, e-mail: sokirinska@gmail.com https://orcid.org/0000-0001-6140-8518	к.е.н., доцент, доцент кафедри економіки і менеджменту, Західнодонбаський інститут ПрАТ « ВНЗ Міжрегіональна Академія управління персоналом»
Roman Savchuk **Роман Савчук**	Senior Instructor of the Department of Economic Informatics, Ukrainian State University of Science and Technologies, Dnipro, Ukraine, Dnipro, Ukraine, e-mail: d.v.savchuk@ust.edu.ua https://orcid.org/0000-0002-3336-6565	ст. викладач кафедри економічної інформатики, Український державний університет науки і технологій, Дніпро, Україна

UDC 004.9, 004.514.62
https://doi.org/10.31891/csit-2022-4-12

NATALYA SOKOLOVA, MARYNA PETRYHA, TETIANA BULANA
Dnipro University of Technology, Dnipro, Ukraine

TIME MANAGEMENT WEB -ORIENTED INFORMATION SYSTEM BASED ON THE MERN TECHNOLOGY STACK

The problem of fleeting time and its effective use has deep roots in human history. Many generations have asked the question of what time is, how it can be controlled, and if it cannot be controlled, how to deal with this enormous influence on human life. Time is the most valuable resource that you have to learn to manage. The more we manage to do, the better the quality of our work and life in general. And in the conditions of digital transformation and acceleration of changes, a clear order sometimes helps to maintain focus on goals, not to deviate from the course, and, ultimately, to be in demand in the labor market.

The main drivers of the time management system are management methods and techniques. Techniques allow different ways to compose a plan for different periods and evaluate priority. The methods, in turn, solve the issue of the implementation of this plan. The implementation tool is no less important, and here information systems have offered themselves very well. The possibility of dynamic planning, scalability, a system of reminders, and a system of rewards - all reflects powerful solutions for the introduction of time management.

This work offers an overview of the process of developing a convenient information system with time management methods. A study of time management methods was conducted. An information model was developed that combines Pomodoro techniques, ABC analysis based on the Eisenhower principle, and Getting Things Done. This model allows you to classify tasks according to the degree of importance, divide them into different groups and lists, assign time for long-term and short-term tasks, and analyze the degree of completion of each task and the work as a whole.

Modeling of the information system was carried out using the unified modeling language UML and its diagrams of precedents, deployment, classes, packages, and components, which simplify the creation of the application logic and reflect the structure of the application. The design of the system was also carried out using the creation of a prototype in the online vector service Figma. The information system is implemented as a client-server web-oriented application based on the MERN technology stack. React is used to develop the web application interface, Node.js, and the Express.js framework for the application server, which implements the creation and request of APIs for communicating with the document-oriented database management system MongoDB using the mongoose module.

Keywords: time management, client-server architecture, MERN technology stack, Eisenhower matrix, Pomodoro technique, ABC analysis technique

НАТАЛІЯ СОКОЛОВА, МАРИНА ПЕТРИГА, ТЕТЯНА БУЛАНА
Національний технічний університет «Дніпровська політехніка»

ВЕБ-ОРІЄНТОВАНА ІНФОРМАЦІЙНА СИСТЕМА ТАЙМ-МЕНЕДЖМЕНТУ НА БАЗІ СТЕКУ ТЕХНОЛОГІЙ MERN

Проблема швидкоплинності часу та його ефективного використання має глибоке коріння в історії людства. Багато поколінь задається питання що таке час, як його можна контролювати, а якщо його неможливо контролювати, як обходиться з цим величезним впливом на людське життя. Час – це найцінніший ресурс, з яким треба вчитися поводитися. Чим більше задуманого ми встигаємо зробити, тим краща якість нашої роботи та життя в цілому. А в умовах цифрової трансформації та прискорення змін, чітке розпорядження часом допомагає зберігати фокус на цілях, не відхилятися від курсу і, зрештою, бути затребуваним на ринку праці.

Головними рушіями системи тайм менеджменту є методи та техніки з управління. Техніки дозволяють різними шляхами компонувати план на різні періоди та оцінювати пріоритет. Методи в свою чергу вирішують питання виконання цього плану. Не менш важливим є інструмент реалізації, і тут дуже добре себе запропоновали інформаційні системи. Можливість динамічного планування, масштабності, система нагадувань, система нагороджень – все це відображає потужні рішень для запровадження тайм менеджменту.

В даній роботі запропоновано огляд процесу розробки зручної інформаційної системи з методами тайм менеджменту. Проведено дослідження методів тайм менеджменту. Розроблена інформаційна модель, яка поєднує техніки Помодоро, АВС-аналізу на основі принципу Ейзенхауера, Getting Things Done. Дана модель дозволяє проводити класифікацію задач за ступенем важливості, розподіляти їх в різні групи та списки, призначати час виконання довгострокових та короткотермінових завдань, проводити аналіз ступеня виконання кожного завдання та роботи в цілому.

Проведено моделювання інформаційної системи уніфікованою мовою моделювання UML та її діаграм прецедентів, розгортання, класів, пакетів, компонентів, які спрощують створення логіки додатку та відображають структуру додатку. Також проведено проектування системи за допомогою створення прототипу у векторному онлайн сервісі Figma. Інформаційна система реалізована у вигляді клієнт-серверного веб-орієнтованого додатку на основі стеку технологій MERN. React використовується для розробки інфтеріейсу веб-додатку, Node.js та фреймворк Express.js для сервера додатку, де реалізовані створення та запит API для комунікування з документо-орієнтованою системою управління базою даних MongoDB за допомогою модуля mongoose.

Ключові слова: тайм-менеджмент, клієнт-серверна архітектура, стек техгологій MERN, матриця Ейзенхауера, техніка Pomodoro, методика АВС-аналізу

Introduction

Back in the days of ancient Rome, the famous thinker and statesman Lucius Annaeus Seneca reflected in his first of the «Moral Letters to Lucilius» on how not to waste time: «Time is the only and most valuable resource you

possess. Be strict with your time, know what you're spending it on, and appreciate when others give you their time.» Also, Seneca kept constant records of time in written form, just as he kept records of money [1].

A person's life and well-being are affected not only by a well-organized professional sphere, but also by the ability to self-organize. If there are opportunities that you don't want to lose, there is a need to realize your desire to have more and more space and responsibility for the ability to manage your plans.

Currently, the pandemic has provided even more free time. According to the Prometheus educational platform, during the pandemic, the number of active students of online courses increased 4 times, but the average rate of people completing the selected courses is only 13% [2, 3]. There are many reasons for this, but good time management can solve the issue of effectively managing your plans for any period and timely analysis of the already traveled path. Time management is a set of tools, methods, processes of rational, conscious control over the amount of time spent on a certain activity, aimed at increasing efficiency and productivity. The term expanded its meaning not only as control of work activities, but also of personal ones. Time management methods have reached various solutions for balancing goals related to family, hobbies, and self-development [4].

Time management allows you to solve various problems of human efficiency and work capacity: self-management, setting goals, focusing on the most important on a certain path, achieving results. Thanks to time management techniques, labor productivity increases, which in turn motivates to do better and achieve the maximum; responsiveness to relevant life challenges increases; it is possible to analyze the importance of mandatory or unnecessary questions [5].

Time management includes such actions as setting a goal, creating a plan, prioritizing, organizing work, decomposing as needed, delegating resources, monitoring and analyzing the path traveled, creating your own work system based on the analysis. The tool is suitable for creating various plans with different terms and substances. These plans can be divided into projects, detailed plans, or a simple list of tasks. In turn, these processes turn into a plan with a to-do list, schedule, calendar [6].

At the same time, time management has many human factors of the inability to correctly reproduce and the consequences of this inability - procrastination: a state when, in the process of work or study, any unexpected small message, not necessarily important, attracts all attention and concentration to itself and distracts from performing important tasks by switching away from unimportant ones, such as scrolling through the news feed, for example. This is solved by self-discipline, which in turn is a reward system or statistics of time spent on useless, unplanned actions [7].

Related works

There are a lot of applications that to some extent implement certain time management methods, for example, the implementation of the full Pomodoro method or the use of only the so-called time-boxing [8], but almost none of them implement time management comprehensively.

Web-based productivity, self-improvement, and habit-enhancing application with the ability to gamify processes and accomplish tasks while staying motivated and having fun at the same time, Habitica [9]. It has functions of to-do list management, task checklist, habit formation, functional Kanban board tracking, healthy and unhealthy habit analysis, which influence the reward system. Disadvantages include the complex appearance of the site, the inability to create your own groups of tasks, the gamification of the reward system can cause game addiction, conditional free, because full functionality can only be obtained through a paid plan.

A program for managing tasks to improve productivity Tweek [10]. This tool is built based on of a weekly calendar view without hourly planning, provides the ability to view weekly tasks, has checklists and subtasks, including recurring tasks, and is synchronized with Google Calendar. The disadvantages are the lack of ability to create lists of groups or folders, analysis of task performance.

MyLifeOrganized [11] is a task, project, and habit management tool designed to balance the simple with the complex. Once the data is loaded, the tool will create a simple list that contains only those actions that require immediate attention. The application has the functions of working with a to-do list, time tracking, mobile access, creating subtasks, but there is no calendar function, and as a drawback, it is worth mentioning the high monthly subscription fee.

Clear [12] is a simple to-do list app that allows you to cross off items on a list, but is only available on the iOS platform.

Purpose

An important factor in successfully mastering skills and using time management with the help of applications is the ease of use of the platform. This is influenced by the adaptability of the Web application and good UI/UX, ease of interaction with the functionality, cross-platform. In addition, for the psychological side of the issue, the ability to analyze usage in the form of statistics and a reward system is important, therefore the development of a convenient time management information system that combines basic time management techniques in its implementation is relevant.

Proposed technique

To implement time management, it is recommended to analyse of time use and planning for each equivalent period based on what has been done before this period and what is planned for the next. Such an opportunity is provided by the technique of prioritizing tasks for time management "ABCD analysis", or analysis according to the Eisenhower principle. According to legend, the matrix was invented by the 34th president of the USA, Dwight Eisenhower, and he used the matrix during his presidency. There was no official confirmation of this. In 1984, the matrix was first described by Lothar Seivert [13]. The technique is based on sorting categories of big data into groups, which uses the principles of importance and urgency [14] (Fig.1).

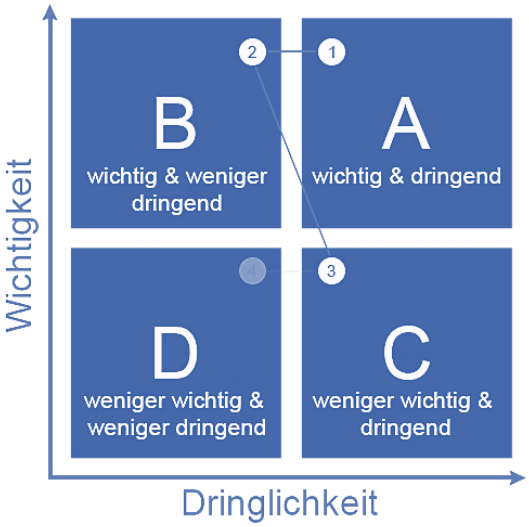


Fig.1. The Eisenhower matrix

Tasks are classified as follows:

a) Type "A" - important and urgent, do not delegate to anyone and do as soon as possible;

b) Type "B" - important and non-urgent, do not delegate to anyone, but set a deadline or delegate to subordinates and leave time for control;

c) Type "C" - unimportant and urgent, must be delegated to subordinates if they do not need to have special knowledge and skills;

d) Type "D" - unimportant and non-urgent, it is advised to abandon the tasks of this group altogether.

The technique of Brian Tracy, a specialist in self-development, is formulated as follows: "eat a frog for breakfast", where the frog is the most difficult task for a certain day, then the rest of the work will be completed almost effortlessly [7].

The Pareto technique [15] is a continuation of Brian Tracy's technique, which is useful when several task directions compete for attention. In the final step of the analysis of the profitability of each task, the one that will bring the overall benefit is chosen. This technique should be limited to the exclusion of important problems at the discretion of the performer, which at first glance and at a specific moment in time are not big, but have a claim to grow over time. This technique is more useful in combination with analytical tools, such as the analysis of the types and consequences of failures, which analyzes and identifies the most critical steps of production processes for the purpose of quality management, and the analysis of the "tree of failures", which with the help of graph edges and logical operators links element failures with object failures. Also, this technique is called the "80/20" rule, where only 20% of all cases provide 80% of the desired result.

The ABC-analysis technique is similar to the Pareto technique. Tasks are divided into three groups, according to their importance, which is based on three laws [16]:

a) Type "A", or the most important cases, which are equal to 15% of the total number of all cases. These cases provide about 65% of the desired result;

b) Type "B", or important tasks, which are equal to 20% of the total number of all cases. In turn, the significance of these cases is 20% of the desired result;

c) Type "C", or unimportant tasks, which are equal to 65% of the total number of cases. Their contribution is 15% to overcome the goal.

It can be seen that the latter techniques involve a third-party solution to the importance of the task. In the case of the Eisenhower principle, at the stage of providing a classification for the "unimportant, urgent" group, it is necessary to additionally determine the degree of preparedness of the subordinate to whom the task will be delegated. The method of determining this argument is at the discretion of the user, which makes each existing case of using these techniques unique and subject to extensive analysis.

There is a combination of ABC-analysis with the Eisenhower principle (Fig. 2). It does not solve the fallibility of determining importance, but it reduces the amount of data for selection and analysis.

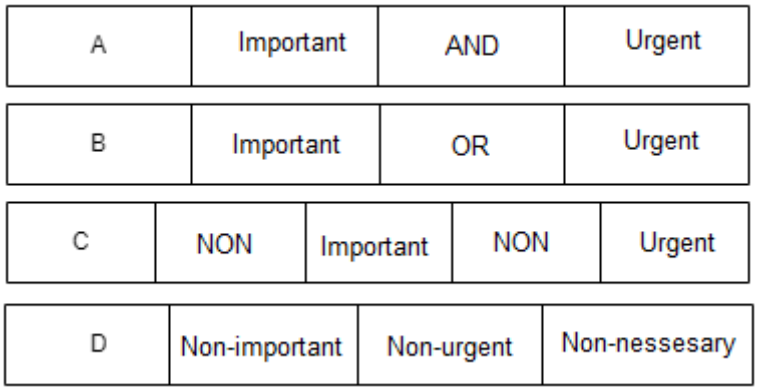


Fig.2. Combination of ABC with the Eisenhower principle

It has the following three type groups:

a) Type "A", or important and urgent. In a real case, no more than one or two such tasks per day. You need to solve such a task or tasks yourself with increased attention;

b) Type "B", a task that is either important and non-urgent, or unimportant and urgent. In this type, the task owner must either make an artificial deadline to complete the task personally, or delegate to a subordinate and leave time for review;

c) Type "C", a task that is not important for various non-business purposes and has an indefinite deadline. Such deliveries must be delegated to a subordinate.

You can ignore the option of type "D", which are unimportant and non-urgent, most often these tasks are not necessary for business purposes or work, but if they bring pleasure or recreation, then they can be kept.

A normal working day or the solution of business tasks should contain or should be performed first only those tasks that have the type "A" and "B", or "A", "B" and "C" [14].

The Pomodoro method is suitable for learning and work [17]. A 30-minute period of time is called a *tomato* - 25 minutes for work and 5 minutes for rest. In its simplest form, it works like this:

- make a list of tasks for the day based on your priorities;
- turn on the timer for 25 minutes, start working;
- after 25 minutes, take a break of 5-10 minutes (warm up or refresh the drink);
- after four tomatoes, take a long break for 15-20 minutes;
- continue to work for 25 minutes. After completing the task, cross it out and move on to the next one.

The Pomodoro method helps you get more done in less time. At the same time, workaholics also use it: they limit themselves in order to rest. When working on a timer, the user knows the value of his time, works more productively, manages expectations, trains willpower and prevents burnout. The timer is called "tomato" (Pomodoro) because Francesco Cirillo originally used a small kitchen timer in the shape of a tomato.

GTD, Getting Things Done, or the technique to make tasks get done, was created by David Allen [5]. The basic idea of the technique is to finish all the small tasks as soon as possible and break the big task into small ones. The reason for looking for this kind of solution was the warning of information overload, when the list of tasks becomes larger every day and is not solved so quickly. The practice of this technique is to put tasks and ideas on paper or not and organize them as quickly as possible so that it is easier to control them.

Results

A time management model has been developed that combines the GTD, Pomodoro and task list methods as follows: it is recommended to create a general task list and a daily task list, which consists of tasks from the general list for a given day. An alternative is to create lists of unimportant tasks. Tasks are prioritized as follows: they are numbered in the order of their importance and are done in this order, they are placed using a technique similar to the ABC method, where type "A" includes tasks for a day, type "B" for a week, type "C" for a month and method prioritization, where A-types include undesirable tasks (Pareto technique) when they are done, others are easier to do, and B-types and C-types can follow the same idea, but instead of necessarily doing exactly an unwanted task, avoiding it will motivate you to do all the other tasks to avoid it.

For the convenience of the user, an implementation in the form of a Web application using the Mern stack was chosen, the key components of which are:

1. MongoDB is a cross-platform, document-oriented database. Data is stored in flexible documents using a special query language based on JavaScript Object Notation.

2. Express. A basic platform designed for creating Web-applications. The platform is an original solution due to its minimalist structure and high speed of operation.

3. React is a JavaScript library designed for organizing user interfaces. The interface uses a functional programming language, not just templates. The main advantage is the execution of the code both in the browser and directly on the server.

4. Node.js is a cross-platform environment used to execute JavaScript. Built on the Chrome V8 JavaScript engine, the platform is used for scalable Web-applications.

The Mern technology stack is a full-stack development, that is, client-side, server-side and database development and the connection between them. The application has a three-level architecture model (Fig. 3).

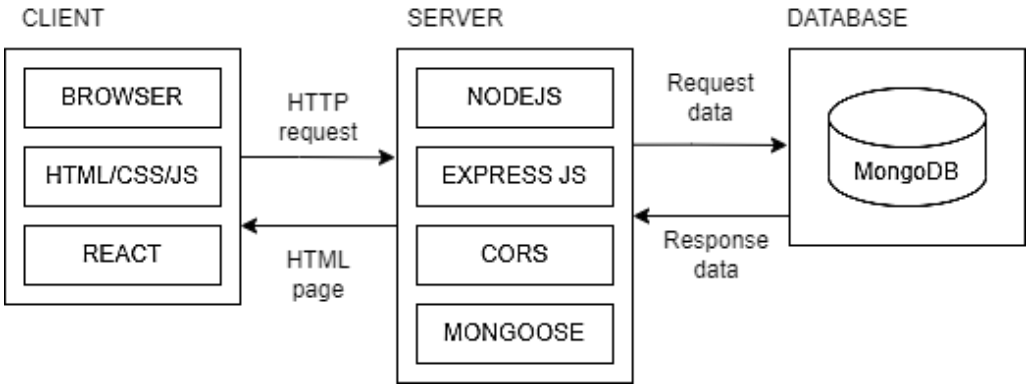


Fig.3. A three-tier application architecture model

The Web-server is the presentation layer and provides the user interface, content is developed using HTML, CSS and Javascript, React. The application server corresponds to the middle layer, which contains the business logic used to process user data. This layer is developed using Node.js and the Express.js framework. A database server is the data layer or server layer of a Web-application. It runs on database management software, in the case of this project and chosen stack, MongoDB. React is used to develop the Web-application's infrastructure, Node.js and the Express.js framework for the application server, which will implement the creation and request API for communicating with the document-oriented MongoDB database management system using the mongoose module.

The structure has a number of advantages, for example, it gives a lot of freedom to update or replace only certain parts of the program without affecting the product as a whole. This allows the application to be extended quite easily by separating the front end from the databases chosen according to the individual needs of the client, while at the same time critical components of the application can be encapsulated and preserved while the whole system continues to evolve organically.

For the developed system, deployment diagrams were constructed to visualize the topology of the physical components of the system and component diagrams to model the physical aspects of the system (Fig.4.). Component diagrams are used to describe components, while deployment diagrams show how they will be deployed in hardware.

The deployment diagram represents the three components of the hardware topology of the system: the client-side server, the Web-server, and the database (Fig.4,a). The component diagram of this project (Fig. 4,b) depicts the client side with the React.js program, the server side with Node.js and the database. React.js is based on a Single Page Application (SPA), and this avoids loading a new page with each action and greatly simplifies the user experience.

The following functions were defined for the system:

- creating a user;
- creation of groups;
- task creation;
- task review;
- assignment of eigenvalues for the Pomodoro timer;
- analysis of satistics.

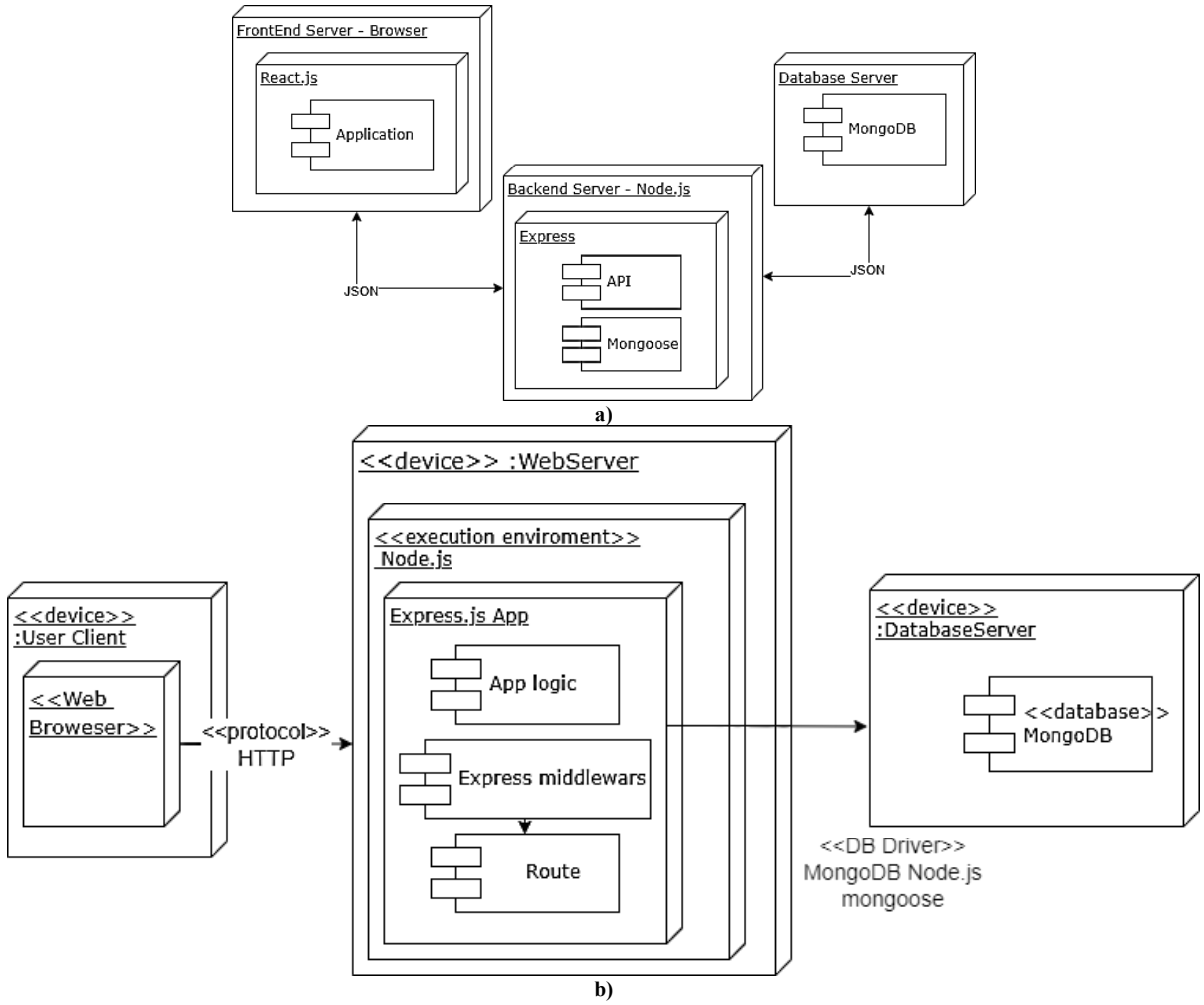


Fig.4. Visualization of the topology of the physical components of the system to model the physical aspects of the system: Component Diagram; b) Deployment Diagram

UML case diagrams, which model system functionality using actors and use cases, were defined for design choices and development priorities. Diagrams of precedents for this project with a list of all possible actions (Fig. 5).

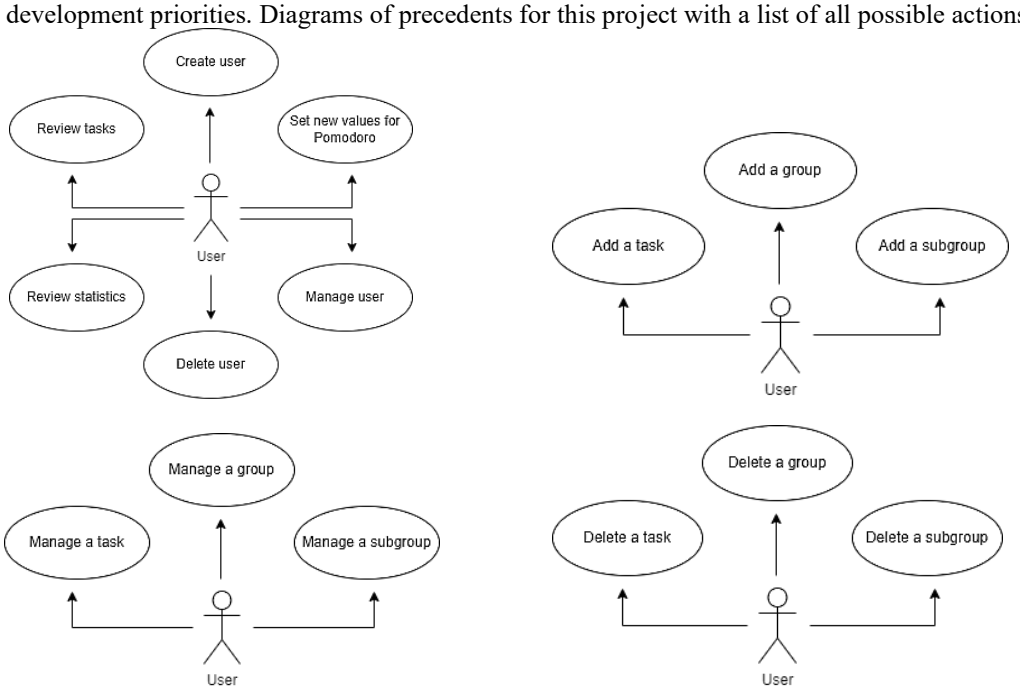


Fig.5. Precedents Diagram

The class diagram represents the static view of the application. It represents the types of objects in the system and the relationships between them. The class diagram of this project has a type of relationship with composition, that is, if a user is deleted from the database, then all groups and tasks created by him are deleted. Also, if a user is created, it can have many groups, or none, but if a subgroup of groups is created, or a task in a subgroup, then there can be from one to many in the parent class (Fig. 6).

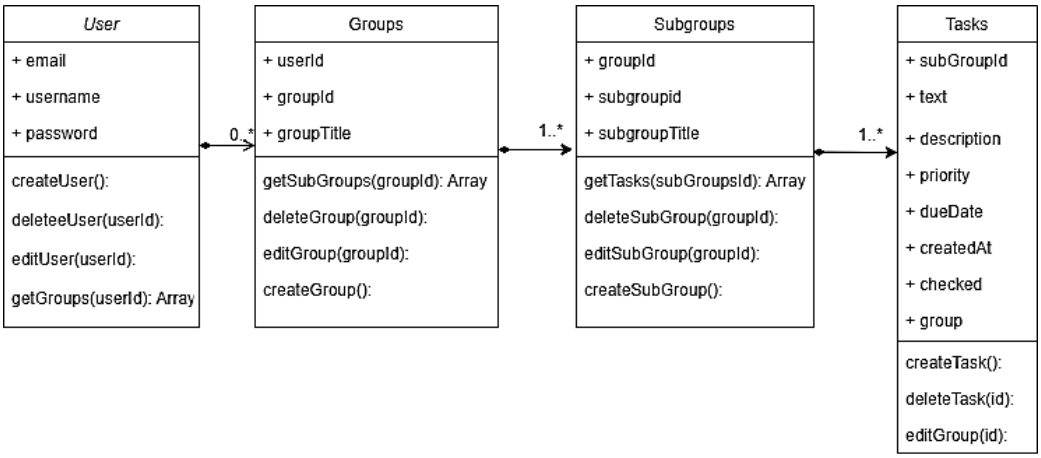


Fig.6. Class diagram

The object diagram reflects the specific (actual) behavior of the system (Fig. 7).

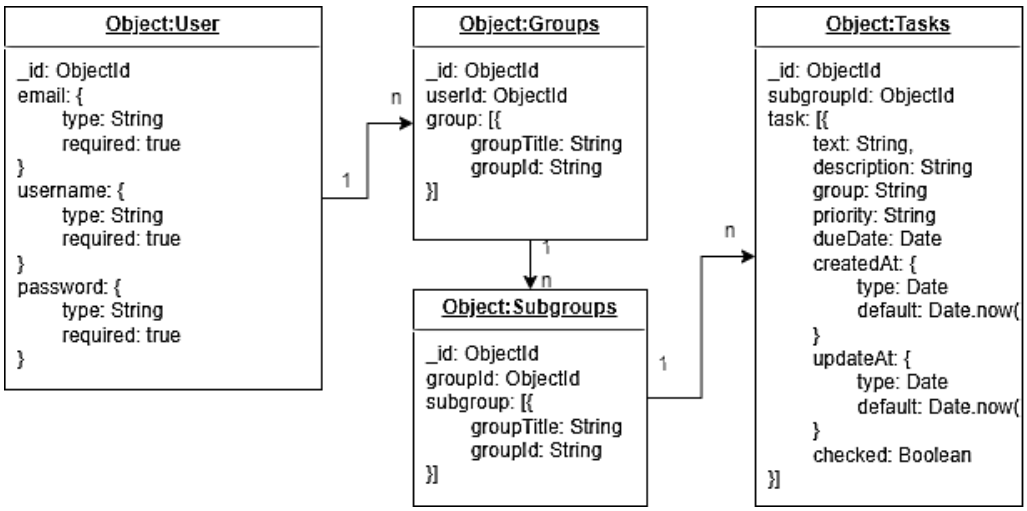


Fig.7. Object Diagram

The server part consists of the following modules (Fig. 8):

- config – file with services, project services;
- routes – is a file with all paths, "routes", with specified callback functions;
- models – designed database tables;
- controllers – file with appropriate callback functions for paths.

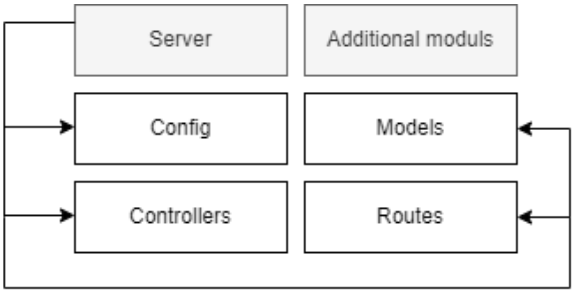


Fig.8. The structure of the server part

The following tools were used to develop the server:

- cors is an HTTP header-based mechanism that allows the server to specify any sources (domain, scheme, or port) other than its own from which the browser should allow resources to be downloaded;
- dotenv is a zero-dependency module that loads environment variables from an .env file into process.env;
- express is a Node.js framework designed to quickly create APIs and facilitate node js work;
- mongoose is a MongoDB object modeling tool designed to work in an asynchronous environment;
- nodemon is a tool that helps develop Node.js applications by automatically restarting the node application when directory file changes are detected;
- jsonweb token is an open standard that defines a compact and autonomous way to securely transfer information between parties as a JSON object;
- bcryptjs is a password hash function;
- body-parser – parses the body of incoming requests in the middleware before the handlers are available in the req.body property.

The client part has the following structure (Fig.9).

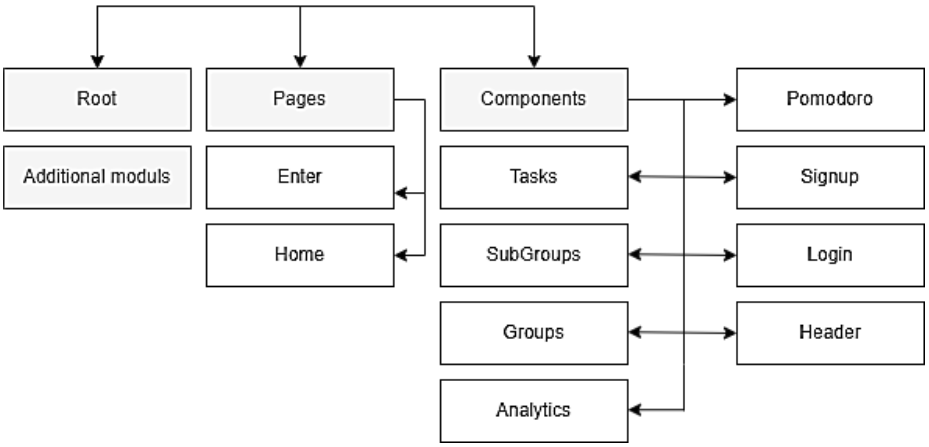


Fig.9. Structure of the client part

A batch diagram, which is a type of structure diagram, shows the location and organization of model elements in a project. This project has the following batch diagram shown on Figure 10.

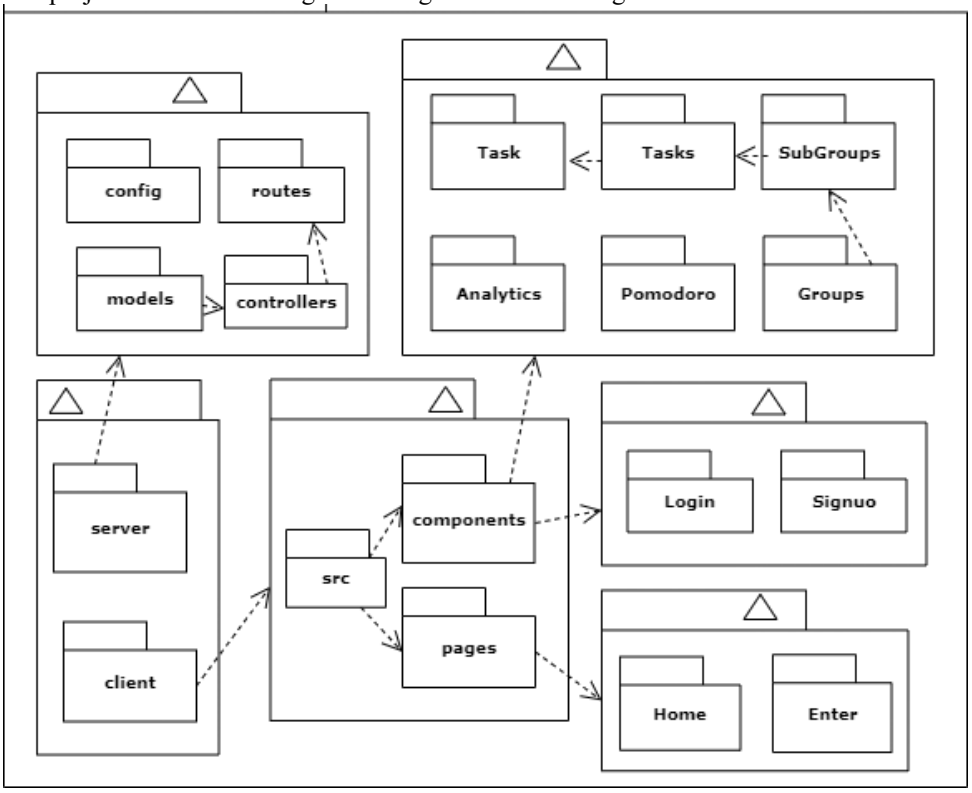


Fig.10. Batch diagram

With the help of the online vector service Figma, a prototype of the Web application was developed, which made it possible to test, simulate and virtually explore the usability of the product before its creation. The application has a minimalistic UI/UX design (Fig.11, 12), easy to use, and has a good emphasis on the components that the user interacts with. By using the React library and implementing a one-page creation method, the user does not waste time loading pages, thanks to the library's component approach. This increases the user experience with the site and its ease of use.

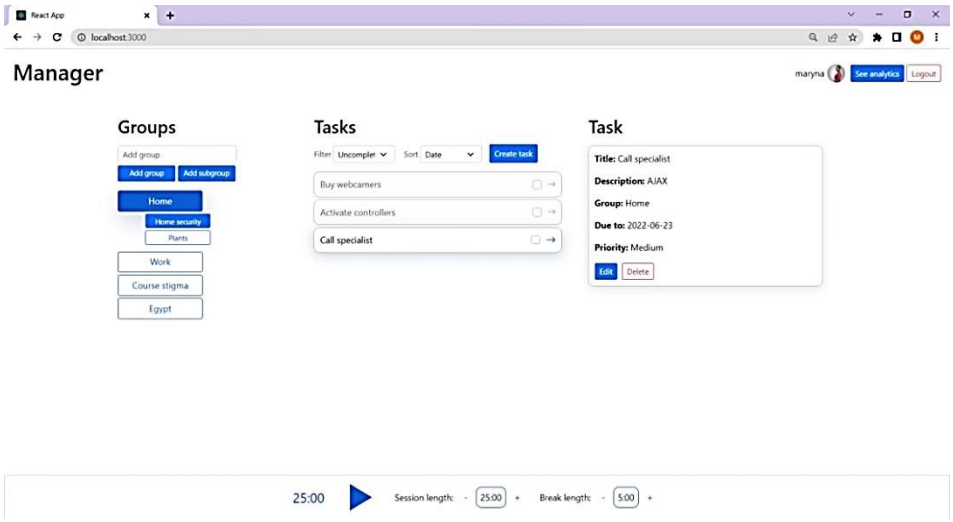
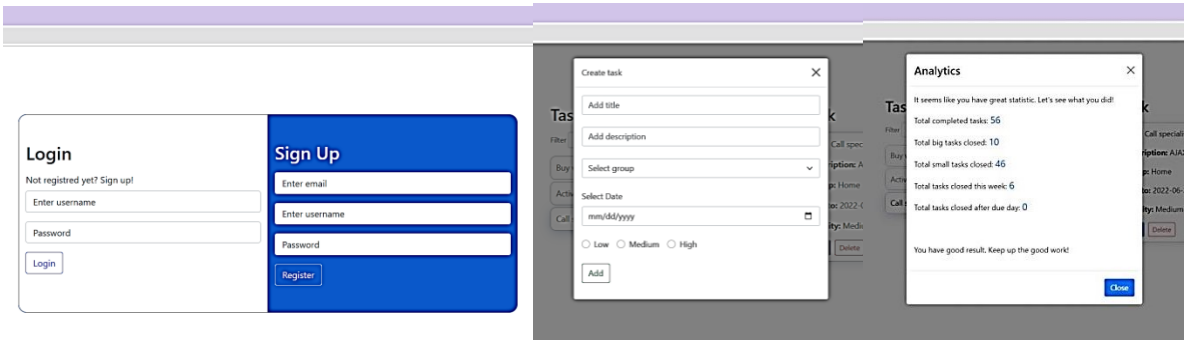


Fig.11. Screenshot of the main page



a) **b)** **c)**

Fig.12. Screenshots: a) the Login screen; b) the task creation window; c) the analysis module window

Conclusions

The work developed a model that combines basic time management methods. Modeling of the information system based on a 3-level architecture based on the MERN technology stack was carried out using the unified modeling language UML and diagrams of precedents, deployment, classes, packages, components, which simplify the creation of the application logic and reflect the structure of the application. The design of the system was also carried out using the creation of a prototype in the online vector service Figma. A Web application with basic time management methods has been created.

As a result of the work, a user-friendly developed Web application was obtained with basic time management methods such as the Pomodoro timer, the ability to divide tasks into groups and subgroups and the priority of tasks and analyze the result with a minimalistic UI/UX design for convenient implementation of time management.

In the process of development, a system of rewards and incentives for completing tasks, which will be integrated into the system in order to motivate the user.

References

1. Thoughts on Moral Letters to Lucilius: Letter 1 - On Saving Time. URL: https://0xedward.io/thoughts-on-moral-letters-to-lucilius-letter-1-on-saving-time/
2. Іван Примаченко про онлайн-навчання на карантині. URL: https://www.radiosvoboda.org/a/30582636.html
3. Social eLearning: A Strategy for Increasing Online Course Completion Rates. URL: https://www.novoed.com/resources/blog/elearning-social-completion-rates/#:~:text=Completion%20rates%20in%20online%20courses,low%2C%20averaging%20around%2013%25.
4. Stella Cottrell. The Study Skills Handbook / Stella Cottrell. – Palgrave Macmillan, 2013. – 123p.
5. David Allen. Getting Things Done: The Art of Stress-Free Productivity, 2015. - 352p.
6. Time Management. URL: https://corporatefinanceinstitute.com/resources/management/time-management-list-tips/
7. Brian Tracy. Eat That Frog!: 21 Great Ways to Stop Procrastinating and Get More Done in Less Time, 2017. - 144p.
8. Pomodoro Technique Considered Harmful URL: https://arialdomartini.wordpress.com/2014/05/19/pomodoro-technique-considered-harmful-dont-worry-you-are-not-using-it/

9. Habitica URL: https://habitica.com/
10. Tweek URL: https://tweek.so/
11. MyLifeOrganized URL: https://www.mylifeorganized.net/
12. Clear URL: https://apps.apple.com/us/app/clear-todos/id493136154
13. Lothar Seiwert. Mehr Zeit für das Wesentliche. So bestimmen Sie Ihre Erfolge selbst durch konsequente Zeitplanung und effektive Arbeitsmethodik. mi (Moderne Industrie), Landsberg am Lech 1984 (Erstausgabe), ISBN 3-478-32770-9.
14. Н.С. Краснокутська, О.М. Нащекіна, О.В. Замула та ін. Менеджмент: навчальний посібник. Харків: «Друкарня Мадрид», 2019. – 231 с.
15. Richard Koch. The 80/20 Principle. Nicholas Brealey Publishing, 1997, Doubleday 1998. ISBN 0-385-49174-3
16. ABC Method Time Management: What It Is & How to Implement. [URL: https://biz30.timedoctor.com/abc-method-time-management/#2-alpen-method
17. Francesco Cirillo. The Pomodoro Technique. URL: https://francescocirillo.com/products/the-pomodoro-technique#

Natalya Sokolova **Наталя Соколова**	PhD, Associate Professor of Information Tecnology & Computer Engineering Department, Dnipro University of Technology, Dnipro, Ukraine email: n.olegowna@gmail.com https://orcid.org/0000-0003-2493-3553	кандидат технічних наук, доцент кафедри Інформаційних технологій та комп'ютерної інженерії, Національний технічний університет «Дніпровська політехніка», Дніпро, Україна
Maryna Petryha **Марина Петрига**	Bachelor of Information Systems and Technologies, Dnipro University of Technology, Dnipro, Ukraine email: maryna.petryha@gmail.com https://orcid.org/0000-0003-0253-7699	бакалавр, Національний технічний університет «Дніпровська політехніка», Дніпро, Україна.
Tetiana Bulana **Тетяна Булана**	PhD, Associate Professor of Information Tecnology & Computer Engineering Department, Dnipro University of Technology, Dnipro, Ukraine email: tatyana.bulanaya@gmail.com https://orcid.org/0000-0001-6346-3326	кандидат технічних наук, доцент кафедри Інформаційних технологій та комп'ютерної інженерії, Національний технічний університет «Дніпровська політехніка», Дніпро, Україна

UDC 004.9: 004.056.53: 004.622
https://doi.org/10.31891/csit-2022-4-13

ANATOLIY SACHENKO, OLEKSANDR OSOLINSKYI, VOLODYMYR KOCHAN,
OLEH SACHENKO, PAVLO BYKOVYY, DIANA ZAHORODNIA
West Ukrainian National University

CONCEPT OF INTELLIGENT MEASURING SYSTEM FOR ANALYZING THE ENERGY CONSUMPTION OF IOT MODULES

The conducted analysis showed the shortcomings of known techniques for measuring MCU (microprocessor systems and IoT modules - hereinafter MCU) current consumption.

In order to improve self-powered systems, designed to perform complex algorithms (learning ANNs, searching for signatures of suspicious software code, etc.), it is necessary to minimize the energy consumption of software, since hardware of modern MCU has low consumption. Optimizing MCU software in terms of energy consumption requires experimental studies of the energy consumption of the execution of instructions, commands and programs. Due to the lack of specialized measuring equipment that has sufficient accuracy (the error must be less than the difference in power consumption of instructions and commands) and immunity to interference, known models do not allow optimizing the software in terms of power consumption and thereby increasing the operating time of autonomously powered systems without recovery battery charge. And this is important for MCU and IoT modules operating in remote locations, dangerous for people, and critical infrastructure systems. However, a direct study of the energy consumption of MCU instructions, commands and programs as part of IoT modules is very time-consuming. RISC microcontrollers have few commands, but many modes of their execution, so it would be necessary to investigate 10-20 thousand options. It is proposed to use the methods of artificial intelligence for the classification of teams and the forecast of energy consumption of those teams that were not studied.

The basis of the concept of a measuring system for the analysis of current consumption of smart devices and IoT modules can be the architecture developed in this paper. At the same time, a more detailed study of them and provision of noise protection, in particular, due to non-contact measuring devices, is required.

Keywords: self-powered system, complex algorithm, software power consumption.

АНАТОЛІЙ САЧЕНКО, ОЛЕКСАНДР ОСОЛІНСЬКИЙ, ВОЛОДИМИР КОЧАН,
ОЛЕГ САЧЕНКО, ПАВЛО БИКОВИЙ, ДІАНА ЗАГОРОДНЯ
Західноукраїнський національний університет

КОНЦЕПЦІЯ ВИМІРЮВАЛЬНОЇ СИСТЕМИ ДЛЯ АНАЛІЗУ СПОЖИВАННЯ СТРУМУ РОЗУМНИМИ ПРИСТРОЯМИ ТА МОДУЛЯМИ ІОТ

Проведений аналіз показав недоліки відомих методик вимірювання струму споживання мікроконтролера.

З метою вдосконалення систем з автономним живленням (мікропроцесорних систем та IoT модулів – надалі МК), призначених для виконання складних алгоритмів (навчання ШНМ, пошук сигнатур підозрілого програмного коду та ін.), необхідно мінімізувати енергоспоживання програмного забезпечення (ПЗ), оскільки апаратне забезпечення (АЗ) сучасних МК має мале споживання. Оптимізація ПЗ МК за енергоспоживанням вимагає експериментальних досліджень енергоспоживання виконання інструкцій, команд і програм. У зв'язку із відсутністю спеціалізованого вимірювального устаткування, яке має достатню точність (похибка має бути меншою за різницю енергоспоживання інструкцій та команд) і завадостійкість, відомі моделі не дають змоги оптимізувати ПЗ за енергоспоживанням і збільшити тим самим час роботи систем з автономним живленням без відновлення заряду акумуляторів. А це важливо для МК та IoT модулів, що працюють у віддалених місцях, небезпечних для людей, та систем критичних інфраструктур. Однак пряме дослідження енергоспоживання інструкцій, команд і програм МК в складі IoT модулів дуже трудомістке. RISC мікроконтролери мають мало команд, але багато режимів їх виконання, тому треба було би дослідити 10 – 20 тисяч варіантів. Пропонується використати методи штучного інтелекту для класифікації команд і прогнозу енергоспоживання тих команд, які не досліджувалися.

Основою концепції вимірювальної системи для аналізу споживання струму смарт-пристроями та IoT-модулями є розроблена у статті архітектура. При цьому потрібне більш детальне забезпечення шумозахисту, зокрема, за рахунок безконтактних вимірювальних приладів.

Ключові слова: система з автономним живленням, складний алгоритм, енергоспоживання програмного забезпечення.

Introduction

The main reason for difficulties in measuring the power consumption of MCU is their current consumption as peaks synchronous with clock pulses. The reason for this is the CMOS technology of their production. The average current consumption of the MCU was measured using a shunt and a digital voltmeter [1-9]. At the same time, uncontrolled energy exchange led to significant errors. If the capacity of the capacitor in the power supply circuit is 1000 μF, a change in voltage on it by 10 μV corresponds to the energy of executing 25 instructions of MCU ARM7TDMI [1-6]. Turning on the shunt between the capacitor and MCU introduces parasitic inductance [7, 8]. Even 0.1 μH at a clock frequency of 25 megahertz creates a resistance of 15 ohms. At a peak current of 50 milliamps, this creates a voltage drop of 0.75 V. But at least the fifth harmonic, that is, a frequency of 125 megahertz, must be taken into account. Therefore, such switching on of the shunt leads to MCU failures. The scheme of current measurement using an operational amplifier with current consumption in the form of peaks causes MCU failures due to the limited response speed of the amplifier [9]. The scheme for measuring the instantaneous current consumption of MCU using

a current mirror [10] has a drawback - the effect of voltage changes on the base-emitter junction of transistors with large current changes. MCU does not go astray. But these changes lead to changes in the voltage on the MCU up to 0.5 - 0.6 V. This creates a methodical error in the consumption current measurement, which is difficult to correct. Therefore, the error of the models built according to the methods [10] reaches 7-10%.

The scientific idea of the authors is to combine the method of measuring the instantaneous power consumption of the microcontroller and the method of measuring the average power consumption of the MCU, as well as deep neural networks for predicting the energy of instructions that will not be studied, which will ensure high accuracy of the measurement results, their mutual comparison and addition. Very often there is a situation when new IoT modules controlled by a microcontroller are being researched and self-adaptation of the system to new types of processor systems made by CMOS technology is necessary. An organic combination of such methods allows you to use the advantages of each of them.

Currently, the main technology of microelectronics is CMOS. The IoT industry is rapidly developing and the basis of all IoT devices are embedded microprocessor systems (microcontroller-based systems) with autonomous power. Their components, manufactured using CMOS technology, have a high speed of execution of instructions with low energy consumption. However, to date, the problem of extending the time of their autonomous operation without recharging the batteries is a constant one. An obvious method of increasing the autonomous operation time of MCU is to increase the energy capacity of power sources. This requires fundamental research on the creation of new materials and their testing. The second way is to improve CMOS technologies, which is also time-consuming. The third way is the optimization of the hardware structure, which is actively used. But it applies only to new developments. Optimizing software by energy consumption is a promising and universal way. However, optimization of software by energy consumption requires the development of a mathematical model of the energy consumption of the processor core. Such a model should be based on the results of experimental studies. At the same time, existing models have low accuracy (7-10%). To increase the accuracy, it is necessary to use equipment that ensures high accuracy of measuring the power consumption of CMOS microcircuits, in which the current is consumed at the moments of change in the logical state of the elements. It should be noted that today various devices are manufactured using CMOS technology - microprocessors, microcontrollers, specialized microcircuits, memory, logic elements, programmable logic matrices. The nature of their energy consumption is determined only by CMOS technology, and the methods and tools proposed in this project are suitable for all these devices. However, the currently existing methods and means of measuring the current, power, and energy of software execution have a large error, because they do not take into account the pulse nature of the energy consumption of devices made according to CMOS technology. The devices made as part of the preparation for the defense of dissertations by A. Borovyi and O. Osolinskyi were separate models. They made it possible to check the possibility of achieving high accuracy of measuring the impulse energy consumption of MCU. To ensure the possibility of developing optimal software, it is necessary to create a specialized intelligent system that combines the above-mentioned systems in a way that makes it possible to automate the process of researching the energy consumption of various types of MCU.

Conceptual model of the measuring system

Based on the analysis of the methods that were described in the previous section, several characteristics and requirements were identified that must be met by the method of measuring the current strength in the MCU power supply circuit:

1. Non-contact method of measuring the current and voltage in the MCU power supply circuit, the contact method is permissible only with theoretical and experimental confirmation that the noise level is permissible;
2. Sufficiently small measurement error of current and voltage in the MCU power supply circuit, as well as a bit rate sufficient to analyze the results of changes in the executable code;
3. The ability to filter out external interference that can affect the result of measuring current and voltage;
4. Possibility (if necessary) to amplify the received signal;
5. Control of the output current and voltage of the power source of the tested module/MCU;
6. The ability to control the MCU clock generator in a wide frequency range;
7. Necessity of mutual censorship of measurement results and rejection of false results.
8. The possibility of accumulating and classifying large volumes of data that will come from measuring channels;
9. Possibility of using neural networks for Matched filter functions.

Based on this, the system architecture can be synthesized (Fig. 1).

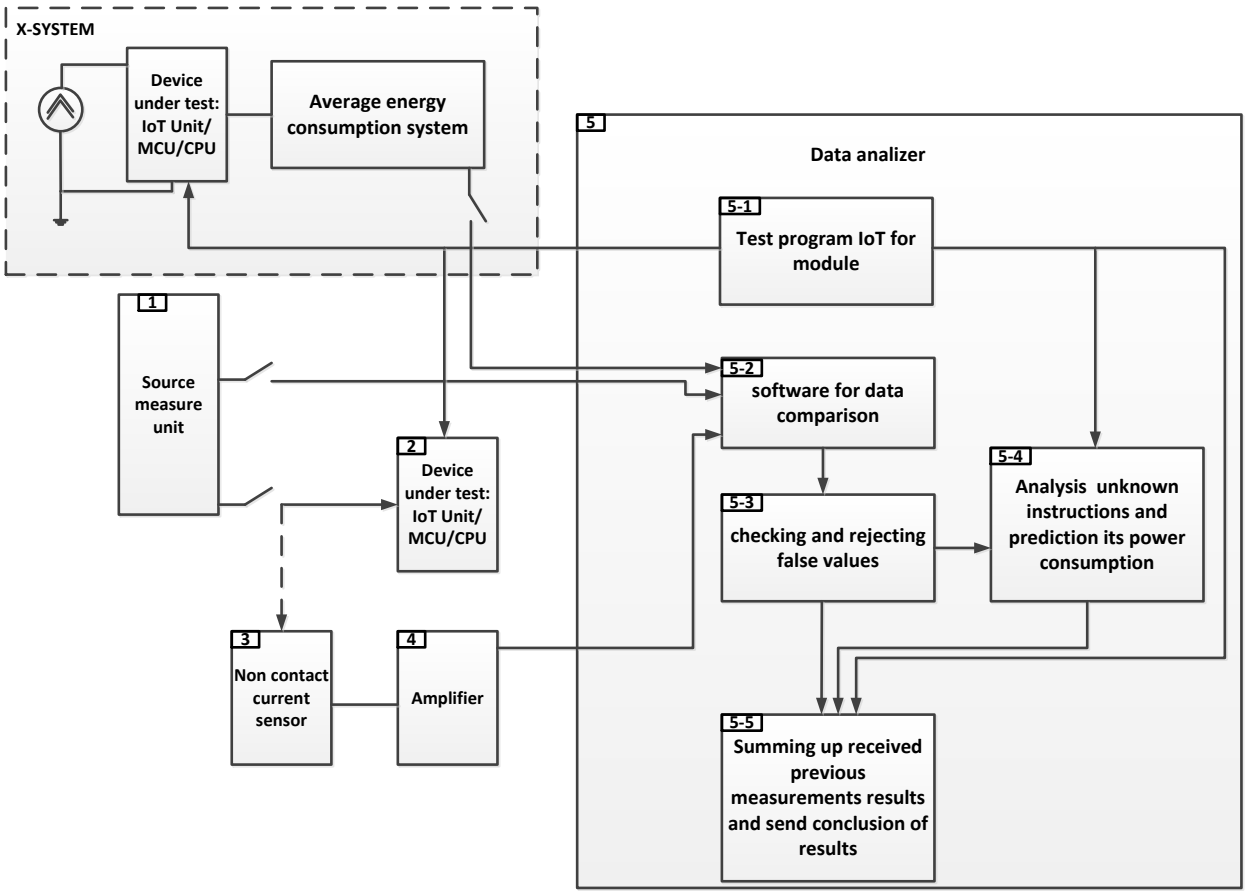


Fig. 1 Architecture of the intelligent system for measuring the energy consumption of IoT module

The system includes:

1. Source measure unit (SMU) – a specialized current source with the ability to measure its output voltage and current with sufficient accuracy and resolution;
2. Device under test (DUT) – a researched smart device or IoT module;
3. Current Sensor - Non contact current sensor - in this case, it is a non-contact device for current-current conversion
4. Amplifier - signal amplifier (current or current-voltage converter);
5. Data Analyzer – A set of software tools (software) for processing measurement results from unit 1 and 4

The Data Analyzer itself includes:

a) The system for downloading the executable code for the IoT module, that is, special software for MK where you can create code, compile it and download it for execution.
b) Software for obtaining, accumulating and comparing measurement results
c) A specialized program for censoring the received data from block 5.2 and rejecting false values in the measurement database
d) A software module based on deep learning ANNs for parallel analysis of the boot code for the IoT module and the energy consumption evaluation results of this code, if the data on the energy consumption of unknown instructions is included in the array of measurement results, then this software predicts the probable energy of executing the instructions or code fragment.
e) The software module summarizes the results about the total energy consumption of the executed code and displays detailed information about code improvements or critical blocks of code where there is the largest share of energy consumption.

The system can also include additional energy consumption analysis systems to improve the measurement results, for example, the MK average energy consumption measurement system (X-SYSTEM) [26] This is due to the fact that the system is divided into hardware and software parts, and blocks 5b) and 5d) above can take several vectors of the array for measuring data.

Thus, the essence of the idea is a universal inexpensive measuring tool (system), and additional elements can be added to increase the accuracy due to the tasks that are set before the developers of software and hardware of IoT modules.

Conclusions

When analyzing the methods and techniques for measuring the current consumption of MCU and IoT modules, or other MCU-based devices, it was found that this issue requires a more in-depth comprehensive approach to measurements. The conducted analysis showed the shortcomings of known techniques for measuring MCU current consumption. The task is relevant because there are many MCU-based devices on the market, especially smart devices or IoT devices, and a set of methods and tools is needed to evaluate their performance and the possible effectiveness of optimizing the executable code.

Thus, the basis of the concept of a measuring system for the analysis of current consumption of smart devices and IoT modules can be the methods proposed in [24-26] and the architecture developed above. At the same time, a more detailed study of them and provision of noise protection, in particular, due to non-contact measuring devices, is required.

References

1. T. Karnik. Technology trends, requirements and challenges for ubiquitous self-powered IOT systems deployment. 2018 Ninth International Green and Sustainable Computing Conference (IGSC), 2018. Pp. 1-1. https://doi.org/10.1109/IGCC.2018.8752113.
2. R. Salvador, A. Sanchez, X. Fan and T. Gemmeke. A Cortex-M3 Based MCU Featuring AVS with 34nW Static Power, 15.3pJ/inst. Active Energy, and 16% Power Variation Across Process and Temperature. 2018 - IEEE 44th European Solid State Circuits Conference (ESSCIRC), 2018. Pp. 278-281. https://doi.org/10.1109/ESSCIRC.2018.8494312.
3. S. Persia et al. IoT Enabling Technologies for Extreme Connectivity Smart Grid Applications. 2019 CTTE-FITCE: Smart Cities & Information and Communication Technology (CTTE-FITCE), 2019. Pp. 1-6. https://doi.org/10.1109/CTTE-FITCE.2019.8894819.
4. T. Mathanlal, J. Martin-Torres, A. Bhardwaj and M. Z. Mier. Self-Sustainable Monitoring Station for Extreme Environments (S3ME2): Design and validation. 2018 Second International Conference on Green Computing and Internet of Things (ICGCIoT), 2018. Pp. 240-245. https://doi.org/10.1109/ICGCIoT.2018.8753046.
5. R. Bannatyne. Microcontrollers can Tough it out in Extreme Environments. URL: https://www.electropages.com/blog/2016/10/microcontrollers-can-tough-it-out-in-extreme-environments
6. N. Mohod. Role of Microcontroller in Automation. 2018. http://dx.doi.org/10.13140/RG.2.2.29448.24328
7. Y. Shi, G. Qiu, & N. Wang. Basic Microcontroller Use for Measurement and Control. In Holden, N. M., Wolfe, M. L., Ogejo, J. A., & Cummins, E. J. (Ed.), Introduction to Biosystems Engineering. 2020. https://doi.org/10.21061/IntroBiosystemsEngineering/Microcontroller
8. MCU Basic Structure/Operation. URL: https://www.renesas.com/sg/en/support/engineer-school/mcu-01-basic-structure-operation
9. M. Chui, M. Collins, and M. Patel. IoT value set to accelerate through 2030: Where and how to capture it. URL: https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/iot-value-set-to-accelerate-through-2030-where-and-how-to-capture-it
10. G. Nick. How Many IoT Devices Are There in 2022? [All You Need To Know]. URL: https://techjury.net/blog/how-many-iot-devices-are-there/#gref
11. Guide to reduce the Arduino Power Consumption. URL: https://diyi0t.com/arduino-reduce-power-consumption/
12. CHAPTER 2 PIC® Microcontroller Low Power Tips ‘n Tricks. TIP #5 Reduce Operating Voltage. URL: https://ww1.microchip.com/downloads/en/DeviceDoc/01146B_chapter%202.pdf
13. F. Reverter and M. Gasulla. Experimental Characterization of the Energy Consumption of ADC Embedded into Microcontrollers Operating in Low Power. 2019 IEEE International Instrumentation and Measurement Technology Conference (I2MTC), 2019. Pp. 1-5. https://doi.org/10.1109/I2MTC.2019.8826815.
14. Successive-approximation ADC. URL: https://en.wikipedia.org/wiki/Successive-approximation_ADC
15. C. W. Lai, S. Ziegler and H. H. C. Iu. High gain amplification of low-side current sensing shunt resistor. 2008 Australasian Universities Power Engineering Conference, 2008. Pp. 1-5.
16. Low-Side Current Sense Circuit Integration. URL: https://www.ti.com/lit/an/sboa190/sboa190.pdf?ts=1652467959102&ref_url=https%253A%252F%252Fwww.google.com%252F
17. What Is a Source Measurement Unit or SMU? URL: https://www.analog.com/en/analog-dialogue/studentzone/studentzone-december-2017.html
18. L. Chiou, C. Yang and C. Chang. A Data-Traffic Aware Dynamic Power Management for General-Purpose Graphics Processing Units. 2019 IEEE International Symposium on Circuits and Systems (ISCAS), 2019. Pp. 1-5. https://doi.org/10.1109/ISCAS.2019.8702405.
19. I. K. Huzmiev and Z. A. Chipirov. Energy consumption powered by graphics processing units (GPU) in response to the number of operating computing unit. 2016 2nd International Conference on Industrial Engineering, Applications and Manufacturing (ICIEAM), 2016. Pp. 1-4. ttps://doi.org/10.1109/ICIEAM.2016.7910995.
20. B. Dimitrov, G. Collier, A. Cruden. Design and Experimental Verification of Voltage Measurement Circuits Based on Linear Optocouplers with Galvanic Isolation for Battery Management Systems. World Electr.Veh. J. 2019. Vol. 10. Article 59. https://doi.org/10.3390/wevj10040059
21. Digital Isolator ADUM1100. URL: https://www.analog.com/en/products/adum1100.html
22. M. Beshley, K. Przystupa, O. Kochan, B. Rusyn, R. Stanisławski, I. Kahalo (2022). 5G multi-tier radio access network planning based on voronoi diagram. Measurement. 2022. Vol. 192. Article 110814. https://doi.org/10.1016/j.measurement.2022.110814
23. C. Wang, H. Jiao, L. Anatychuk, N. Pasyechnikova et al. (2022). Development of a Temperature and Heat Flux Measurement System Based on Microcontroller and its Application in Ophthalmology. Measurement Science Review. 2022. Vol. 22. Issue 2. Pp. 73-79. https://doi.org/10.2478/msr-2022-0009
24. O. Osolinskyi, V. Kochan, G. Sapozhnyk, A. Sachenko and O. Kochan. Stand for investigating the measuring methodical errors of microcontrollers average energy power consumption. 2018 International Conference on Development and Application Systems (DAS), 2018. Pp. 105-109. https://doi.org/10.1109/DAAS.2018.8396080.
25. O. Oleksandr, A. Molga, V. Kochan and A. Sachenko. Method of Ensuring the Noise Immunity at Measurement of Single-Board Microcontroller Average Energy Within IoT Environment. 2020 IEEE 40th International Conference on Electronics and Nanotechnology (ELNANO), 2020. Pp. 807-810. https://doi.org/10.1109/ELNANO50318.2020.9088880.
26. O. R. Osolinskiy. Advanced system measuring energy consumption of microcontrollers. Measuring and Computing Devices in Technological Processes. 2015. Vol. 3. Pp. 104-108. (in Ukrainian)

Anatoliy Sachenko **Анатолій Саченко**	DrSc (Engineering), Professor, Professor of Department for Information-Computing Systems and Control, West Ukrainian National University https://orcid.org/0000-0002-0907-3682 e-mail: as@wunu.edu.ua	Доктор технічних наук, професор, професор кафедри інформаційно-обчислювальних систем і управління, Західноукраїнський національний університет
Oleksandr Osolinskyi **Олександр Осолінський**	CSc (Engineering), Associated Professor, Associated Professor of Department for Information-Computing Systems and Control, West Ukrainian National University https://orcid.org/0000-0002-0136-395X e-mail: oso@wunu.edu.ua	Кандидат технічних наук, доцент, доцент кафедри інформаційно-обчислювальних систем і управління, Західноукраїнський національний університет
Volodymyr Kochan **Володимир Кочан**	CSc (Engineering), Professor, Professor of Department for Information-Computing Systems and Control, West Ukrainian National University https://orcid.org/0000-0001-8376-4660 e-mail: vk@wunu.edu.ua	Кандидат технічних наук, професор, професор кафедри інформаційно-обчислювальних систем і управління, Західноукраїнський національний університет
Oleh Sachenko **Олег Саченко**	CSc (Engineering), Associated Professor of Department for Specialized Computer Systems, West Ukrainian National University https://orcid.org/0000-0001-9337-8341 e-mail: os@wunu.edu.ua	Кандидат технічних наук, доцент кафедри спеціалізованих комп'ютерних систем, Західноукраїнський національний університет
Pavlo Bykovyy **Павло Биковий**	CSc (Engineering), Associated Professor, Associated Professor of Department for Information-Computing Systems and Control, West Ukrainian National University https://orcid.org/0000-0002-5705-5702 e-mail: pb@wunu.edu.ua	Кандидат технічних наук, доцент, доцент кафедри інформаційно-обчислювальних систем і управління, Західноукраїнський національний університет
Diana Zahorodnia **Діана Загородня**	CSc (Engineering), Associated Professor of Department for Information-Computing Systems and Control, West Ukrainian National University https://orcid.org/0000-0002-9764-3672 e-mail: dza@wunu.edu.ua	Кандидат технічних наук, доцент кафедри інформаційно-обчислювальних систем і управління, Західноукраїнський національний університет

UDC 519.7:007:004
https://doi.org/10.31891/csit-2022-4-14

YURII TURBAL
The National University of Water and Environmental Engineering
SERHII BABYCH
Rivne Vocational College of NUBiP

INFORMATION TECHNOLOGY FOR THE SCHEDULE GENERATION BASED ON THE ALGEBRA OF ADDITIVE-DISJUNCTIVE FORMS AND THE MODIFIED METHOD OF PERMANENT DECOMPOSITION

To improve the information technology of drawing up class schedules, there is a need to develop methods that allow significantly reduce the number of combinatorial objects in the process of algorithms for generating schedules matrices. For example, the result of applying the method of permanent decomposition is a collection of combinatorial objects - permutations, combinations, and placements. For the task of drawing up lesson schedules in the part of forming timetable matrices, the method provides a memory-recorded set of all possible systems of various representatives of sets, which are the columns of the timetable matrices (SRPS). Since the algorithm of permanent decomposition gives all possible SRPS, it creates the problem of forming the final schedule based on SRPS or all possible variants of schedules and requires the development of special algorithms. Certain known approaches to solving such a problem are associated with significant computational complexity in the general case. This also applies to the approach based on the order relation of the set of decomposition matrices.

The basis of the information technology proposed in the work is the further modification of the incidence matrices and, accordingly, such a modification of the permanent decomposition method, which allows generating ready versions of the schedule matrices at the output. This is achieved due to the introduction of a special algebra of additive-disjunctive forms and, accordingly, the possibility of generating such forms in the process of permanent decomposition. In fact, in this context, ADF is a formal representation of a ready-made version of an admissible schedule that satisfies some additional requirements.

Keywords: information technology, additive-disjunctive form, permanent, decomposition.

ЮРІЙ ТУРБАЛ
Національний університет водного господарства та природокористування
СЕРГІЙ БАБИЧ
ВСП Рівненський фаховий коледж НУБіП України

ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ СКЛАДАННЯ РОЗКЛАДУ НА ОСНОВІ АЛГЕБРИ АДДИТИВНО-ДИЗ'ЮНКТИВНИХ ФОРМ ТА МОДИФІКОВАНОГО МЕТОДУ ПЕРАМНЕНТНОЇ ДЕКОМПОЗИЦІЇ

З метою вдосконалення інформаційних технологій складання розкладів занять виникає необхідність у розробці методів, що дозволяють суттєво скоротити кількість комбінаторних об'єктів в процесі роботи алгоритмів генерації матриць розкладів. Наприклад, результатом застосування методу перманентної декомпозиції є сукупність комбінаторних об'єктів – перестановок, комбінацій, розміщень. Для задачі складання розкладів занять в частині формування матриць розкладів метод дає записану в пам'ять сукупність усіх можливих систем різних представників множин, що являють собою стовпчики матриць розкладів (СРПС). Оскільки алгоритм перманентної декомпозиції дає всі можливі СРПС, то це породжує проблему формування остаточного розкладу на основі СРПС чи усіх можливих варіантів розкладів та вимагає розробки спеціальних алгоритмів. Окремі відомі підходи до розв'язання такої задачі пов'язані з значною обчислювальною складністю в загальному випадку. Це стосується і підходу на основі відношення порядку на множині матриць розкладу.

В основі інформаційної технології, що пропонується в роботі, покладена подальша модифікація матриць інцидентності та, відповідно, така модифікація методу перманентної декомпозиції, яка дозволяє на виході генерувати готові варіанти матриць розкладів. Це досягається за рахунок введення спеціальної алгебри аддидитво-диз'юнктивних форм та відповідно, можливість генерації таких форм в процесі перманентної декомпозиції. По суті, в такому контексті АДФ є фактично формальним представленням готового варіанту допустимого розкладу, який задовольняє низку додаткових вимог.

Розроблена інформаційна технологыя вирішення задач календарного планування, зокрема задачі формування розкладів, є цілісною системою, що поєднує деякі підходи, зокрема конфігураційний підхід до аналізу вхідних даних та алгоритмів формування вихідних допустимих матриць розкладів, застосування алгоритмів перманентної декомпозиції, лексикографічний підхід на основі відповідних порядкових відношень, алгебру адитивно-диз'юнктивних форм.

Ключові слова: інформаційна технологія, аддитивно-диз'юнктивна форма, перманент, декомпозиція.

Introduction

A significant amount of most recent research has focused on the tasks of scheduling. There are numerous excellent resources available in the cloud. The cost of performing tasks in the cloud depends on what resources are used. Cloud planning is different from traditional planning. In the environment of cloud computing, the task of scheduling is the biggest and most difficult issue. The task scheduling problem is the NP-complete problem. Many heuristics have introduced scheduling algorithms, but more improvements are needed to make the system faster and more responsive. [1]

Although a significant number of algorithms have been developed to generate various combinatorial objects, such as permutations, permutations with repetitions of different types, and systems of subsets of some sets of elements [2-4, 5-12], new approaches and algorithms are still emerging.

Given the novelty of the combination of permanent and decomposition solutions within the calendar calculation - it is difficult to rely on similar literature.

Many different task scheduling problems such as assignment, job-shop, flow-shop, vehicle routing, and other scheduling problems have been studied intensively. The studied grid task scheduling problem in this work comes from the task-resource assignment problem [13] which is much more complicated than the above-stated classic task scheduling problems. Restated, a grid application is a task scheduling problem involving partially ordered tasks and distributed heterogeneous resources, and can be represented by a directed acyclic graph (DAG) [14–17]. The scheduling target is to find optimal task-resource assignment and hence minimize application completion time. Most scheduling problems are confirmed to be NP-complete. Thus, many researchers have devoted their efforts to solving task scheduling problems. The exact algorithm such as branch-and-bound method is able to find the optima of the scheduling problem. However, the execution time required is impractical as the number of tasks and resources increases. Hence, many different schemes have been presented for solving scheduling problems. Chen et al. [18] combined a competitive scheme with slack neurons into Hopfield neural networks to solve multiprocessor real-time job scheduling problems. Sandnes [19] presented a stochastic approach of employing randomization in the scheduling of tasks in multiobjective scheduling problems. An artificial immune-system based scheme was proposed to solve the dynamic economic dispatch problem of generating units [20]. Comparatively, several metaheuristics such as genetic algorithm (GA), simulated annealing algorithm (SA), tabu search (TS), ant colony optimization (ACO), and the particle swarm optimization have been effectively proposed for solving these difficult problems. Oh and Wu [21] presented a multiobjective genetic algorithm, which aims to minimize the number of processors required and the total tardiness of tasks. Liu and Wang [22] solved the resource-constrained project scheduling problem of minimizing activities' cost based on GA. And a thermal generating unit's commitment scheduling problem was studied by a modified GA [23]. Tabu search is an approach to prevent the search from trapping into the local minimum, and it has been applied to solving a single machine scheduling problem with distinct due windows to minimize total weighted earliness and tardiness [27] as well as job-shop scheduling [28].

A detailed overview of the combinatorial algorithms can be given by Knuth [3], and Ruskey [4] which considers the concept of combinatorial generation and distinguishes the following tasks: listing–generating elements of a given combinatorial set sequentially, ranking – numbering elements of a given combinatorial set, unranking - generating elements of a given combinatorial set per their ranks and random selection– generating elements of a given combinatorial set in random order.

General methods for developing combinatorial generation algorithms were studied by such researchers as S. Bacchelli [1], V.V. Kruchinin [5, 6], P. Flajolet [7] and others. It is a well-known algorithm for permutation generation [39], such as Bottom-Up, Lexicography, Johnson-Trotter [40], PIndex [41], and Inversion [42].

Additive-disjunction forms and permanent decomposition method

Suppose we have n elements $(a_1, a_2, \dots, a_m)$, that can be part of m sets $(W_1, W_2, \dots, W_m)$, and the occurrence of the same element several times is allowed. Information about which elements are included in the corresponding sets will be given in the form of an incidence matrix of the form:

$$\begin{pmatrix} & a_1 & a_2 & \dots & a_n \\ W_1 & n_{11} & n_{12} & \dots & n_{1n} \\ W_2 & n_{21} & n_{22} & \dots & n_{2n} \\ \dots & \dots & \dots & \dots & \dots \\ W_m & n_{m1} & n_{m2} & \dots & n_{mn} \end{pmatrix} \quad (1)$$

The elements $(a_1, a_2, \dots, a_m)$ will be called identifiers of the columns of the incidence matrix. The system of different representatives (SDR) will be called a vector of the form:

$$(v_1, v_2, \dots, v_m), v_i \in W_i, i = \overline{1, m}, v_i \neq v_j, i \neq j.$$

We divide identifier elements into the regular and "stream". If the element a_i is "stream", then it must be simultaneously written in all positions of the sample vector, where the correspondent incidence matrix column contains non-zero elements. An arbitrary vector of samples (or a matrix, the rows of which are samples) will be called a schedule.

The schedule

$$((v_{11}, v_{12}, \dots, v_{1m}), (v_{21}, v_{22}, \dots, v_{2m}), \dots, (v_{k1}, v_{k2}, \dots, v_{km}))$$

will be considered correct under the conditions:

1.$\forall j \in \{1,2,\dots,m\}: \{v_{1j} \cup v_{2j} \cup \dots \cup v_{kj}\} = \{n_{j1} * a_1 \cup n_{j2} * a_2 \cup \dots \cup n_{jn} * a_n\}$,

$l * a = \{a_1, a_2, \dots, a_l\}, a_i = a, i = \overline{1, l}$.

2. $\forall i \in \{1,2,\dots,k\}: v_{ij} \neq v_{ir}, j \neq r.$,elements v_{ij}, v_{ir}, are non-stream.

Obviously, in the case when each element is included in each set only once and all elements are non-stream, the matrix of incidence is

$$\begin{pmatrix} & a_1 & a_2 & \dots & a_n \\ W_1 & 1 & 1 & \dots & 1 \\ W_2 & 1 & 1 & \dots & 1 \\ \dots & \dots & \dots & \dots & \dots \\ W_n & 1 & 1 & \dots & 1 \end{pmatrix} \quad (2)$$

Then one of the variants of the correct schedule can be written in the form:

$$\begin{pmatrix} a_1 & a_2 & a_3 & \dots & a_n \\ a_2 & a_3 & a_4 & \dots & a_1 \\ a_3 & a_4 & a_5 & \dots & a_2 \\ \dots & \dots & \dots & \dots & \dots \\ a_n & a_1 & a_2 & \dots & a_{n-1} \end{pmatrix}. \quad (3)$$

The rows of the schedule matrix consist of n permutations of the corresponding elements.

Let us have an arbitrary matrix of the daily timetable, dimension mx n. Consider the columns of the schedule matrix $(R_1, R_2, \dots R_n)$. As already discussed above, the modified incidence matrix in the presence of flows was constructed as follows. Teachers' numbers are displayed horizontally, and groups in which classes are held vertically. Each teacher is assigned a column of the matrix in which zeros and ones are recorded depending on whether the teacher has pairs in the corresponding groups. Moreover, if a teacher has a current pair, then we allocate a separate column of the incidence matrix to him, marking it (you can use an index, which is the number of flow elements). Thus, a matrix of the form is:

$$A = \begin{array}{c} \\ R_1 \\ R_2 \\ \dots \\ R_m \end{array} \begin{array}{c} \begin{array}{cccc} x_1 & x_2 & \dots & x_n \end{array} \\ \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} \end{array}, \quad (4)$$

where $\alpha_{ij} = \begin{cases} 1, & when\ x_i \in R_j, \\ 0 \text{ in another case.} \end{cases}$.

Note that such an incidence matrix does not provide comprehensive information for building a schedule. After all, in the case when the teacher has several pairs in the same group or several similar current pairs, the matrix will still be 1. Therefore, we will consider a modification of the incidence matrix containing information about the number of pairs. For this, instead of 1, we indicate the corresponding number. Therefore

$$\alpha_{ij} = \begin{cases} k_i, & \text{when } x_i \in R_j, k_i - \text{numbers of the lessons} \\ & \text{else } 0. \end{cases}$$

Let's, for example, have a timetable matrix of the form:

$$R = \begin{pmatrix} a & a & a \\ b & c & a \\ a & c & d \end{pmatrix} \quad (5)$$

where a,b,c,d-teachers identifiers. It is necessary to build all possible variants of correct SDR.

Construct the incidence matrix :

$$\begin{pmatrix} & a & a^n & b & c & d \\ R_1 & 1 & 1 & 1 & 0 & 0 \\ R_2 & 0 & 1 & 0 & 2 & 0 \\ R_3 & 1 & 1 & 0 & 0 & 1 \end{pmatrix} \quad (6)$$

Let's construct the scheduling process of the modified permanent with "memorization" [1] according to the first line:

$$per\mathrm{mod}_3^{1,2,3}\begin{pmatrix} a & a^n & b & c & d \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 2 & 0 \\ 1 & 1 & 0 & 0 & 1 \end{pmatrix} = 1_1^a * per\mathrm{mod}_3^{2,3}\begin{pmatrix} b & c & d \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix} + 1_1^{a^p} * 1 + 1_1^b * per\mathrm{mod}_3^{2,3}\begin{pmatrix} a & c & d \\ 0 & 2 & 0 \\ 1 & 0 & 1 \end{pmatrix} = 1_1^a *$$

$$(2_2^c per\mathrm{mod}_3 \binom{bd}{01}) + 1_1^{a^n} * 1 + 1_1^b * (2_2^c per\mathrm{mod}_3 \binom{ad}{11}) = 1_1^a 2_2^c 1_3^d + 1_1^{a^n} + 1_1^b 2_2^c 1_3^a + 1_1^b 2_2^c 1_3^d.$$

We get all possible variants of correct SDR: acd, aaa, bca, bcd.

Let's consider in more detail the procedure for scheduling a permanent. Note that a non-zero element in the row of the matrix in the first step means the mandatory presence of the corresponding component in the schedule. In the process of calculating the permanent, we simply use the addition operation +. However, from the point of view of the construction of the schedule, the logic of the addition operation here is completely different, it denotes a "disjunction", the possibility of choosing one of the options of the SRPS. If an element is multiplied, the value of which is greater than 1, then a situation of inclusion of several components is possible. For example in an expression $2_2^c per\mathrm{mod}_3 \binom{ab}{11} = 2_2^c * 1_3^a + 2_2^c * 1_3^b$ addition means the obligatory inclusion of both components, since element 3 is included twice, the element of the incidence matrix is equal to 2. At the same time, in the expression $1_2^c per\mathrm{mod}_3 \binom{ab}{11} = 1_2^c * 1_3^a + 1_2^c * 1_3^b$ adding means choosing one of two options, since element 3 is used only 1 time. We will mark the selection operation with an icon∨. In addition, when using the mandatory inclusion of the

corresponding element of the column to which the non-zero element of the incidence matrix corresponds, we will reduce the value of the element of the incidence matrix by 1. thus, the expressions in the calculation of the permanent can be interpreted not as ordinary sums, but as special operations applied to systems of different representatives or their fragments. In this case, we will no longer have the numerical value of the algebraic constant, but we will get all possible variants of timetables.

Lets consider examples;

$$1_2^d per\mathrm{mod}_3 \begin{pmatrix} a & b & c \\ 1 & 1 & 1 \end{pmatrix} = 1_2^d * 1_3^a + 1_2^d * 1_3^b + 1_2^d * 1_3^c = ADF = 1_2^d * 1_3^a \vee 1_2^d * 1_3^b \vee 1_2^d * 1_3^c$$

$$2_2^d per\mathrm{mod}_3 \begin{pmatrix} a & b & c \\ 1 & 1 & 1 \end{pmatrix} = 2_2^d * 1_3^a + 2_2^d * 1_3^b + 2_2^d * 1_3^c = ADF =$$

$$= (1_2^d * 1_3^a + 1_2^d * 1_3^b) \vee (1_2^d * 1_3^a + 1_2^d * 1_3^c) \vee (1_2^d * 1_3^b + 1_2^d * 1_3^c)$$

Last example demonstrates an important property: in the case when the value of the multiplier element is less than the number of non-zero elements of the expansion line, we haveC_n^kvariants of interpretation of the initial amount in the form of ADF.

Thus, ADF is an expression that includes two operations: + mandatory inclusion and binary selection∨.

It is easy to determine the properties of the algebra of additive-disjunctive forms:

The following properties are obvious:

1. $A \vee A = A$
2. $A \vee B = B \vee A$
3. $A + B = B + A$
4. $A + A = \{A, A\}$
5. $A + B \vee C = A + (B \vee C)$
6. $A + B \vee C = (A + B) \vee (A + C)$

Using these properties, it is possible to convert to multiple systems of various representative configurations:

$$a + b \vee c + d \vee e = (a + b) \vee (a + c) + d \vee e = (a + b + d \vee e) \vee (a + c + d \vee e)$$
$$= (a + b + d) \vee (a + b + e) \vee (a + c + d) \vee (a + c + e).$$

Now consider the previous example from the point of view of ADF. We can modify the permanent schedule procedure to obtain the correct ADF. To do this, at each iteration of the decomposition of the permanent, we will analyze the occurrence of the same elements in all components of the mandatory inclusion. The condition must be fulfilled - the total number of occurrences of the element in the mandatory inclusion block must be equal to its index. If it turns out that this condition is violated, then the mandatory inclusion block will be considered incorrect and must be removed. We have:

$$per\mathrm{mod}_{1,2,3} \begin{pmatrix} a & a^n & b & c & d \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 2 & 0 \\ 1 & 1 & 0 & 0 & 1 \end{pmatrix} = 1_1^a * per\mathrm{mod}_{2,3} \begin{pmatrix} b & c & d \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix} + 1_1^{a^p} * 1 + 1_1^b * per\mathrm{mod}_{2,3} \begin{pmatrix} a & c & d \\ 0 & 2 & 0 \\ 1 & 0 & 1 \end{pmatrix} = 1_1^a *$$

$$(2_2^c per\mathrm{mod}_3 \begin{pmatrix} b & d \\ 0 & 1 \end{pmatrix}) + 1_1^{a^n} * 1 + 1_1^b * \left(2_2^c per\mathrm{mod}_3 \begin{pmatrix} a & d \\ 1 & 1 \end{pmatrix} \right) =$$

$$= \mathrm{ADF} = 1_1^a * \left(1_2^c per\mathrm{mod}_3 \begin{pmatrix} b & d \\ 0 & 1 \end{pmatrix} \right) + 1_1^{a^n} * 1 + 1_1^b * \left(1_2^c per\mathrm{mod}_3 \begin{pmatrix} a & d \\ 1 & 1 \end{pmatrix} \right) = 1_1^a 1_2^c 1_3^d + 1_1^{a^n} + 1_1^b 1_2^c 1_3^a \vee$$
$$1_1^b 1_2^c 1_3^d =$$
$$(1_1^a 1_2^c 1_3^d + 1_1^{a^n} + 1_1^b 1_2^c 1_3^a) \vee (1_1^a 1_2^c 1_3^d + 1_1^{a^n} + 1_1^b 1_2^c 1_3^d).$$

Thus, we get two variants of the timetables:

$\begin{pmatrix} a & c & d \\ a & a & a \\ b & c & a \end{pmatrix}$ and $\begin{pmatrix} a & c & d \\ a & a & a \\ b & c & d \end{pmatrix}$.

Second variant is incorrect because of 1_3^d is used twice.

Taking this into account, we can introduce a procedure for throwing out incorrect situations in mandatory inclusion blocks:

$$1_1^a 1_2^c 1_3^d + 1_1^{a^n} + 1_1^b 1_2^c 1_3^a \vee 1_1^b 1_2^c 1_3^d = 1_1^a 1_2^c 1_3^d + 1_1^{a^n} + 1_1^b 1_2^c 1_3^a \vee 1_1^b 1_2^c 0_3^d$$
$$(1_1^a 1_2^c 1_3^d + 1_1^{a^n} + 1_1^b 1_2^c 1_3^a) \vee (1_1^a 1_2^c 1_3^d + 1_1^{a^n} + 1_1^b 1_2^c 0_3^d) = (1_1^a 1_2^c 1_3^d + 1_1^{a^n} + 1_1^b 1_2^c 1_3^a.$$

In this example, in the process of building the ADF, the property is taken into account that if the element of the schedule is already used in the expression of mandatory inclusion and it is present in other expressions of mandatory inclusion, then its incident value in the following expressions is reduced by 1.

Thus, the modification of the procedure of the schedule of the permanent incidence matrix by constructing the ADF allows obtaining all possible correct options of schedules. The proposed procedure not only solves the problem of the coincidence of schedule elements in rows but also immediately obtains the required number of identical elements in columns by the input data recorded in the modified incidence matrix.

Note that in the absence of situations when the teacher can have several pairs in one group in the process of constructing the ADF, we obtain the ADF of the form (DF)+(DF)+...+(DF), where DF (disjunctive form)=$1_1^{i_1} * 1_2^{i_2} * \ldots * 1_n^{i_n} \vee 1_1^{j_1} * 1_2^{j_2} * \ldots * 1_n^{j_n} \vee \ldots$

In the case when several pairs are allowed in one group or streams, we will get additive constructions of mandatory inclusion (DAF)+(DAF)+...+(DAF), where DAF (disjunction of additive forms) contains constructions of the type

$$1_1^{i_1} * 1_2^{i_2} * \ldots * 1_n^{i_n} \vee 1_1^{j_1} * 1_2^{j_2} * \ldots * 1_n^{j_n} \vee (1_1^{j_1} * 1_2^{j_2} * \ldots * 1_n^{j_n} \vee \ldots) + (1_1^{j_1} * 1_2^{j_2} * \ldots * 1_n^{j_n} \vee \ldots) \ldots$$

The construction of ADF will be carried out based on the set of SRPS. Indeed, if the value of the incidence index of some element in the SRPS is greater than 1 and equal to k, then in this case we will have C_l^k variants of interpretation of the initial amount in the form of ADF. Moreover, any combination of SRPS with k elements will be recorded with a mandatory inclusion operation, between which there will be a selection operation. Such a procedure will make it possible to construct an ADF based on the set of all SRPS.

First, the operation of mandatory inclusion is placed between the groups of SRPS formed as a result of the schedule according to the first row of the incidence matrix. In more detail, mandatory inclusion groups are all SRPs in which the first element is the element to which the non-zero elements of the first row of the incidence matrix correspond. If the index of the element is 1, then all disjunctions will be in the group, if not, then we will have disjunctions of all possible combinations of conjunctive blocks of mandatory inclusion. For example, if the index is 2, we will have initial groups of mandatory inclusion. Next, we will review all the SRPS sequentially by the first elements (the current element is considered one), the second, etc.

If the first element in any group has an index of 1, then it is obvious that all the SRPS in the group is used with the selection operation. If its index is larger, then we will have all possible combinations of mandatory inclusion blocks connected by the selection operation. Note that the result of such a schedule is the sum of disjunctive forms, which is guaranteed to include all the necessary elements of the first column. Next, we consider all SRPS for the second element. The logic of further actions is to ensure the guaranteed achievement of the required number of inclusions of the second element.

For this, we consider the second row of the incidence matrix and all its non-zero elements. Let's consider the first element. We determine which SRPS has it. Then we define the possible options so that it is entered depending on the index. To do this, we number occurrences, generate combinations, and check the required number of occurrences, taking into account the operations of mandatory inclusions. Since there are only disjunctions in each group, with an index greater than 1, it is necessary to choose from several groups (from one only in case of coincidence). Then you can select the groups from which the corresponding SRPS is selected and then it is selected as the only one. At the same time, all other SRPS that do not contain our element is selected from other groups. We put a disjunction between the new groups. Thus, we get a disjunctive form of groups, each of which has corresponding additive constructions of mandatory inclusion.

Suppose that its index is greater. Note that the SRPS with this element can be in several groups of mandatory inclusion (see example). If it stands in only one group, where selection operations are everywhere, it is impossible to make a schedule. If it is in several different groups, then only 1 SRPS can be selected from each such group. If there are index elements and it is possible to form a schedule with the required number of occurrences of this element, then combinations with SRPS with mandatory inclusion operations are formed. At the same time, SRPS mandatory inclusion will appear! Next, we consider the third element, the fourth, etc. General algorithm was considered in [1].

Information technology

The results described above allow us to detail information technology for solving a wide range of calendar planning problems, which is based on the use of algorithms based on decomposition procedures of specially modified permanent incidence matrices.

This technology is a set of methods and approaches, which include:

• methods of input data analysis;

• methods of generating various combinatorial objects - permutations, systems of various column representatives, combinations with a number of additional conditions, which are based on specially developed permanent decomposition algorithms;

• methods of forming all possible admissible variants of matrices

• schedules, including algorithmic solutions, which are based, in particular, on the use of the algebra of additive-disjunctive forms specially developed for such problems;

• methods of presenting data during the operation of generation algorithms, which are adapted to these algorithms and take into account their specificities as much as possible.

Thus, this information technology is a complete system that allows you to solve the tasks of creating schedules in the presence of a number of additional conditions (See Fig.1).

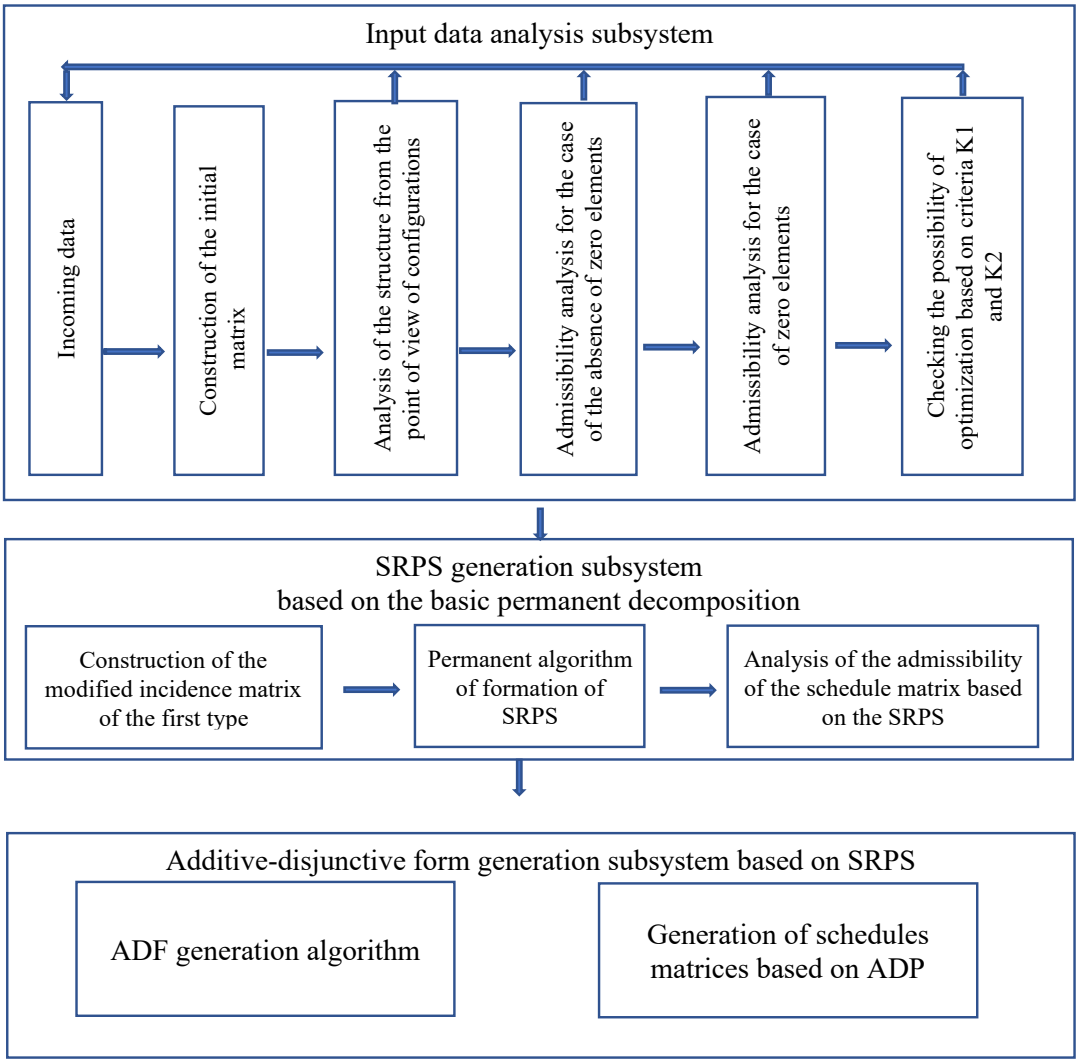


Fig. 1. Information technology based on ADF algebra (a variant of using the basic permanent approach)

Conclusions

The second modification of the incidence matrix is proposed, which differs in that, unlike the previous options, it contains complete information necessary for drawing up a schedule. The corresponding modification made it possible to make significant corrections in the decomposition procedure of the permanent of this matrix and led to the need to develop an algebra of additive-disjunctive forms.

A special algebra of additive-disjunctive forms (ADF) is proposed, which contains two operations - disjunction and mandatory inclusion (addition) and differs in the algorithmic nature of its operations. The operation of disjunction means the selection and, accordingly, the duplication of the corresponding lists containing disjunctive forms in the process of recursive generation procedures. Mandatory inclusion operation means simple inclusion of the corresponding SRPS as the next row in the running schedule matrix.

Based on ADF, two algorithms for the formation of schedule matrices were proposed for the first time:

- the first algorithm allows forming the schedule matrices directly in the process of decomposition of the permanent of the second modification of the incidence matrix;
- the second algorithm allows you to form ADF based on the SMPR, which is formed as a result of the algorithms formulated in the previous section.

This approach is very convenient from the point of view of software implementation and construction of the appropriate class hierarchy. It allows you to create information technology as a complete system that combines appropriate algorithms and methods. Information technology for solving calendar planning problems, in particular, the task of generating schedules, is formulated for the first time, which is a complete system that combines some approaches, in particular, a configuration approach to the analysis of input data and algorithms for the formation of

initial admissible matrices of schedules, the permanent approach of generation of SRPS and application of algorithms of permanent decomposition, a lexicographic approach based on relevant order relations, algebra of additive-disjunctive forms.

References

1. S. Bacchelli, L. Ferrari, R. Pinzani, R. Sprugnoli Mixed succession rules: The commutative case. Journal of Combinatorial Theory. 2010, Vol. 117, Series A. P. 568–582. DOI: 10.1016/j.jcta.2009.11.005

2. Arnaw Wadhonkar A Task Scheduling Algorithm Based on Task Length and Deadline in Cloud Computing / Arnav Wadhonkar, Deepti Theng // International Journal of Scientific & Engineering Research. – 2016. – Vol. 7, № 4. – P. 1905-1909.

3. Knuth D.E. The Art of Computer Programming / D. E. Knuth. – Vol. 4A: Combinatorial Algorithms Part 1. Boston: Addison-Wesley Professional, 2011.

4. Ruskey F. Combinatorial Generation. Working Version (1j-CSC 425/520) [Electronic resourse] / F. Ruskey, Department of Computer Science University of Victoria, 2003. Access mode: http://page.math.tu-berlin.de/~felsner/SemWS17-18/Ruskey-Comb-Gen.pdf. Accessed 1 May 2020.

5. Kruchinin V.V. Methods for Developing Algorithms for Ranking and Unranking Combinatorial Objects Based on AND/OR Trees / V.V. Kruchinin. – Tomsk: V-Spektr, 2007.

6. Shablya Y. Method for Developing Combinatorial Generation Algorithms Based on AND/OR Trees and Its Application / Y. Shablya, D. Kruchinin, V. Kruchinin // Mathematics. – 2020. – Vol. 8, № 962. DOI: 10.3390/math8060962

7. Flajolet P. A calculus for the random generation of combinatorial structures / P. Flajolet, P. Zimmerman, B. Cutsem // Theoretical Computer Science. – 1994. – Vol. 132. – P. 1–35.

8. Optimal algorithms for online scheduling with bounded rearrangement at the end / Chen Xin, Lan Yan, Benkő Attila, Dósa György, Han Xin // Theoretical Computer Science. -2011. – Vol. 412, № 45. – P. 6269–6278. DOI: 10.1016/j.tcs.2011.07.014

9. Do P.T. Exhaustive generation for permutations avoiding (colored) regular sets of patterns // P.T. Do, T.T.H Tran, V. Vajnovszki // Discrete Applied Mathematics. – 2019. – Vol. 268. – P. 44–53.

10. Mirshekarian Sadegh Correlation of job-shop scheduling problem features with scheduling efficiency // Mirshekarian Sadegh, N. Šormaz Dušan // Expert Systems with Applications. – 2016. – Vol. 62. – P. 131–147. DOI: 10.1016/j.eswa.2016.06.014

11. Humble Travis S. Application of Quantum Annealing to Nurse Scheduling Problem / S. Humble Travis, Nakamura Yuma, Ikeda Kazuki. // Scientific Reports. – 2019. – Vol. 9, № 1. – P. 12837. DOI: 10.1038/s41598-019-49172-3.

12. . Henning, K. Jansen, M. Rau, and L. Schmarje, ‘‘Complexity and inapproximability results for parallel task scheduling and strip packing,’’ in Proc. Int. Comput. Sci. Symp. Cham, Switzerland: Springer, 2018, pp. 169–180.

13. T. Chen, B. Zhang, X. Hao, and Y. Dai, “Task scheduling in grid based on particle swarm optimization,” in Proceedings of the 5th International Symposium on Parallel and Distributed Computing (ISPDC ’06), pp. 238–245, July 2006.

14. A. Salman, I. Ahmad, and S. Al-Madani, “Particle swarm optimization for task assignment problem,” Microprocessors and Microsystems, vol. 26, no. 8, pp. 363–371, 2002.

15. M. Aggarwal, R. D. Kent, and A. Ngom, “Genetic algorithm based scheduler for computational grids,” in Proceedings of the 19th International Symposium on High Performance Computing Systems and Applications (HPCS ’05), pp. 209–215, May 2005.

16. G. Malewicz, A. L. Rosenberg, and M. Yurkewych, “On scheduling complex dags for internet-based computing,” in Proceedings of the 19th IEEE International Parallel and Distributed Processing Symposium (IPDPS ’05), April 2005.

17. L. He, S. A. Jarvis, D. P. Spooner, D. Bacigalupo, G. Tan, and G. R. Nudd, “Mapping DAG-based applications to multiclusters with background workload,” in Proceedings of the IEEE InternationalSymposium on Cluster Computing and the Grid, pp. 855–862, May 2005.

18. R. M. Chen, S. T. Lo, and Y. M. Huang, “Combining competitive scheme with slack neurons to solve
real-time job scheduling problem,” Expert Systems with Applications, vol. 33, no. 1, pp. 75–85, 2007.

19. F. E. Sandnes, “Secure distributed configuration management with randomised scheduling of system administration tasks,” IEICE Transactions on Information and Systems, vol. E86-D, no. 9, pp. 1601–1610, 2003.

20. M. Basu, “Hybridization of artificial immune systems and sequential quadratic programming for dynamic economic dispatch,” Electric Power Components and Systems, vol. 37, no. 9, pp. 1036–1045, 2009.

21. J. Oh and C. Wu, “Genetic-algorithm-based real-time task scheduling with multiple goals,” Journal of Systems and Software, vol. 71, no. 3, pp. 245–258, 2004.

22. Z. Liu and H. Wang, “GA-based resource-constrained project scheduling with the objective of minimizing activities’ cost,” in Proceedings of the International Conference on Intelligent Computing (ICIC’05), vol. 3644 of Lecture Notes in Computer Science, pp. 937–946, August 2005.

23. N. Amjady and A. Shirzadi, “Unit commitment using a new integer coded genetic algorithm,”European Transactions on Electrical Power, vol. 19, no. 8, pp. 1161–1176, 2009.

24. S. G. Ponnambalam, P. Aravindan, and S. V. Rajesh, “Tabu search algorithm for job shop scheduling,” International Journal of Advanced Manufacturing Technology, vol. 16, no. 10, pp. 765–771, 2000.

25. S. T. Lo, R. M. Chen, Y. M. Huang, and C. L. Wu, “Multiprocessor system scheduling with precedence and resource constraints using an enhanced ant colony system,” Expert Systems with Applications, vol.34, no. 3, pp. 2071–2081, 2008.

26. W. J. Gutjahr and M. S. Rauner, “An ACO algorithm for a dynamic regional nurse-scheduling problem in Austria,” Computers and Operations Research, vol. 34, no. 3, pp. 642–666, 2007.

27. F. Zhao, Y. Hong, D. Yu, Y. Yang, and Q. Zhang, “A hybrid particle swarm optimisation algorithm and fuzzy logic for process planning and production scheduling integration in holonic manufacturing systems,” International Journal of Computer Integrated Manufacturing, vol. 23, no. 1, pp. 20–39, 2010.

28. Turbal Y. , Babych S., Bachyshyna L., Kunanets N. and Kovalchuk N.. Modification of the Permanent Decomposition Method for the Meeting Schedule Problem. The 1st International Workshop on Information Technologies: Theoretical and Applied Problems (ITTAP-2021). Ternopil, Ukraine. 2021. P.126-131

29. V. Dzhedzhula, I. Yepifanova and Y. Kravchyk, "Use of the Theory of Fuzzy Sets in Determining the Level of Enterprise Security," 2022 12th International Conference on Advanced Computer Information Technologies (ACIT), Ruzomberok, Slovakia, 2022, pp. 311-315, doi: 10.1109/ACIT54803.2022.9913150.

Yurii Turbal **Юрій Турбал**	Dr Sc (Engineering), Professor, Head of the Computer Science and Applied Mathematics Department, National University of Water and Environmental Engineering https://orcid.org/0000-0002-5727-5334 e-mail: turbaly@gmail.com	Доктор технічних наук, професор, завідувач кафедри комп'ютерних наук та прикладної математики Національного університету водного господарства та природокористування
Serhii Babych **Сергій Бабич**	Teacher of programming disciplines, Rivne professional college of the National University of Life and Environmental Sciences of Ukraine, Rivne, Ukraine	Викладач дисциплін програмування, ВСП Рівненський фаховий коледж Національного університету біоресурсів і природокористування України

UDC 004.94:004.042
https://doi.org/10.31891/csit-2022-4-15

KHRYSTYNA LIPIANINA-HONCHARENKO, ANATOLIY SACHENKO
West Ukrainian National University
VITALIY KULYK
Lviv Tourism Development Center - Municipal Enterprise
RUSLAN SAVCHYSHYN
MagneticOne
OLEKSANDRA PROVOZIN
Lviv Tourism Development Center - Municipal Enterprise
SVITLANA SHCHUR
MagneticOne
LILIIA KURPITA
Lviv Tourism Development Center - Municipal Enterprise

SIMULATION MODEL STRUCTURE OF BUSINESS PROCESSES FOR A PRODUCT BASED ON AURALIZATION TECHNOLOGY

Within the framework of the Auralization of acoustic heritage sites using Augmented and Virtual Reality (AURA) project, the task was to develop the business models for potential products that would use the developed auralized 3D model. Accordingly, the purpose of the article is to analyze typical business models for the specified products with an auralized 3D model, which can further be adapted to each specific use case (marketing scenario).

In the most general terms, a business model is a method of doing business that allows a company to sustain itself, i.e. generate income and revenue. Therefore, the development of a business model is relevant to calculate the possible income or expenses of a future product.

The article analyzes existing business models and determines which ones are most suitable for the task at hand. It is proposed to use the following business models: "Canvas", "Barter" and "Joint Ownership".

The "Canvas" business model is described by the following structure: consumer segment, value proposition, sales channels, customer relationships, product revenues, key resources, key activities, key partners, and cost structure. In addition, this business model proposes the E-commerce, Affiliations, and Additional Opportunities business processes.

The "Barter" business model is based on the use of a social media platform to promote products and services. And in the Shared Ownership business model, investors do not buy the entire asset, but only a part of it. For these business models, the Barter and Shared Ownership business processes are presented.

Based on the business processes the structure of the simulation model was proposed. The study can be extended and adapted to other products. In further research, scenario modeling will be conducted using system dynamics approaches.

Keywords: business model, business process, simulation model, auralized 3D model.

ХРИСТИНА ЛІП'ЯНІНА-ГОНЧАРЕНКО, АНАТОЛІЙ САЧЕНКО
Західноукраїнський національний університет
ВІТАЛІЙ КУЛИК
ЛКП "Центр розвитку туризму м. Львова"
РУСЛАН САВЧИШИН
MagneticOne
ОЛЕКСАНДРА ПРОВОЗІН
ЛКП "Центр розвитку туризму м. Львова"
СВІТЛАНА ЩУР
MagneticOne
ЛІЛІЯ КУРПІТА
ЛКП "Центр розвитку туризму м. Львова"

СТРУКТУРА ІМІТАЦІЙНОЇ МОДЕЛІ БІЗНЕС ПРОЦЕСІВ ДЛЯ ПРОДУКТУ НА ОСНОВІ ТЕХНОЛОГІЇ АУРАЛІЗАЦІЇ

В межах проекту Auralisation of acoustic heritage sites using Augmented and Virtual Reality (AURA) була поставлена задача розробити бізнес-моделі для потенційних продуктів, що будуть використовувати розроблену ауралізовану 3D модель. Відповідно, метою статті проаналізувати типові бізнес-моделі для зазначених продуктів за допомогою ауралізованої 3D-моделі, які надалі можуть бути адаптовані до кожного конкретного випадку використання (маркетингового сценарію).

В найзагальнішому визначенні бізнес-модель – це метод здійснення бізнесу, завдяки якому компанія може забезпечити себе, тобто згенерувати дохід, виручку. Тому розробка бізнес-моделей є актуальною, для прорахування можливих доходів чи видатків майбутнього продукту.

У статті проаналізовано існуючі бізнес-моделі та визначено, які найбільше підходять для поставленої задачі. Зоерема, запропоновано використати наступні бізнес-моделі: "Канва", "Бартер" та "Спільна власність".

Бізнес-модель "Канва" описується структурою: споживчий сегмент, ціннісна пропозиція, канали збуту, взаємовідносини з клієнтами, доходи продукту, ключові ресурси, ключові види діяльності, ключові партнери, структура витрат. Також, у даній бізнес-моделі запропоновано бізнес-процес "E-commerce", "Афілювання" та "Додаткові можливості".

При бізнес-модель "Бартер" основою є використання платформи соціальних мереж для просування продуктів і

послуг. А при бізнес-моделі "Спільна власність" інвестори купують не весь актив в цілому, а лише його частку. Для даних бізнес-моделей представлено бізнес-процес "Бартер" та "Спільна власність". На основі бізнес-процесів запропоновано структуру імітаційної моделі. Результати дослідження можна розширити та адаптувати для інших продуктів.
Ключові слова: бізнес-модель, бізнес-процес, імітаційна модель, ауралізована 3D модель..

Introduction

Auralisation – the technique of creating virtual soundscapes in 3D models to provide the same immersive sound experience as the music performed in the real venue. AURA will explore exciting new opportunities that auralisation opens up for music performing arts and their traditional and new audiences.

Auralization technology is closely related to both Augmented and Virtual Reality, but goes in hand with 3D modeling and focuses on sound reproduction. The essence of the technology is to model and reproduce sound depending on the environment and the location of the listener in it.

Prospects for the use of any technology directly depend on how much it will be in demand by users. Therefore, one of the goals of the project "AURA - Auralisation of Acoustic Heritage Sites Using Augmented and Virtual Reality" [1] is to investigate the potential of using auralized 3D models of concert halls in VR and AR, as well as to understand the possibilities of promotion for different target audiences of the product.

During the research, the following groups of potential users were prioritized:

- experts (theaters, designers, musicians, singers, conductors, actors, general expert public, etc.)
- technicians (architectural designers, engineering and architecture students)
- promoters of cultural events (performing organizer, venue holders, public authorities and their culture policy representatives, marketing agencies)

During the marketing research, 21 demonstrative experiences of use succeeded, which are the subject of a separate publication. The purpose of this article is to analyze typical business models for the specified products with an auralized 3D model, which can further be adapted to each specific use case (marketing scenario).

Although business models are fundamentally linked to technological innovations, its design is essentially separated from technology. A business model can be defined as a system that solves the problem of identifying customers in the process of interacting with their needs and satisfactions, taking into account the monetisation of value [2]. Therefore, a number of researchers, in particular Osterwalder, A., & Pigneur, Y [3] and Demil & Lecocq [4], have proposed the concept of a business model as part of a traditional competitive advantage strategy. At the same time, innovative technologies themselves will affect the capabilities of business models [5]. In this regard, the design of business models for a product with an auralised 3D model is actual. Because it will enable a potential developer to determine possible revenues and expenses.

Let us review related References. Reference [6] proposes a new scenario-based predictive approach that allows assessing and predicting the future behavior of business processes based on system dynamics. Reference [7] focuses on the development of a validation technique for business process that result in a detailed report on the partner's process compatibility assessment.

Bocken et al [8], Gassmann et al [9] and Remane et al [10] consider the classification of business models, using a theoretical and template approach. Based on the analysis of these References, for a product that uses an auralised 3D model, it is reasonable to select the following three business models "Canvas"[3], "Barter" and "Joint Ownership" [9]. Hence, we adapted those models for a product that uses an auralised 3D model, illustrated the adaptation process using the Thinging Machine technology [11], and developed a correspond structure of a business process simulation model, that described below.

Business Models of Product with the Auralised 3d Model

Business model 1 "Canvas"

Despite the emergence of several alternatives, the Kanva business model [3] remains perhaps the most widely used [12, 13]. It is structurally divided into nine blocks [14]: Customer segments, Value proposition, Sales channels, Customer relations, Product revenues (revenue streams), Key resources, Key activities, Key partners, Cost structure.

We have refined each structural block of the Kanva business model in relation to a product that uses an auralised 3D model (Table 1). Taking into account the monetisation aspect, it is advisable to consider blocks 3 and 5 in more detail.

Block 3 shows how users of a product that uses an auralised 3D model receive information about value propositions. In particular, an auralised 3D model can promote its capabilities as a product through online channels (website, social media Facebook, Instagram, Telegram, etc.), forums, fairs, advertising, etc.

As an example of a sales channel, it is advisable to consider the use of the E-commerce business process [15], where traditional goods or services are delivered through Internet channels with reduced overheads, which is relevant for an auralised 3D model. The functionalities that can be implemented by e-commerce systems for a product using an auralised 3D model include the following:

- issuing and selling tickets for performances (orders are stored in a single database),
- self-registration of listeners,
- work with loyalty programmes,
- receiving feedback from listeners.

Table 1

Canvas business model for a product that uses an auralised 3D model

<table>
<tr><td colspan="2">8 Key partners</td><td colspan="2">7 Key activities</td><td colspan="2">2 Value proposition</td><td colspan="2">4 Customer relations</td><td colspan="2">1 Consumer segment</td></tr>
<tr><td colspan="2" rowspan="3">Ministry of Social and Cultural Affairs;
Local self-government bodies;
Artists and music experts;
Enterprises of the tourism industry;
Enterprises, institutions and organisations in the social and cultural sector.</td><td colspan="2">Developing a product using an auralised 3D model;
Bringing to market;
Software support;
Research and development and project activities.</td><td colspan="2" rowspan="3">Increasing the creativity of musical and theatrical performances;
Promotion of cultural events;
Modelling the acoustics of concert halls/</td><td colspan="2">A product that uses an auralised 3D model;
Advertising;
CRM systems.</td><td colspan="2" rowspan="2">Different types of classical music listeners.</td></tr>
<tr><td colspan="2">6 Key resources</td><td colspan="2">3 Sales channels</td></tr>
<tr><td colspan="2">Human capital;
Material and technical base;
Financial resources.</td><td colspan="4">Online advertising;
Forums and fairs.</td></tr>
<tr><td colspan="6">9 Costs.</td><td colspan="4">5 Product revenues</td></tr>
<tr><td colspan="6">Labour costs; Costs of providing services; Expenses for marketing services;</td><td colspan="4">Government allocations;</td></tr>
<tr><td colspan="6">Software maintenance costs; Costs of developing a product using an auralised 3D model; Social security costs; Expenditure on items, equipment and inventory;</td><td colspan="4">Monetisation channels;</td></tr>
<tr><td colspan="6">Expenditure on research and development; Utility and energy costs; Debt service costs; Capital expenditure; External and internal borrowing costs; Payments to shareholders.</td><td colspan="4">Investments.</td></tr>
</table>

For example, let us illustrate the E-commerce business process for a product that uses an auralised 3D model (Fig. 1).

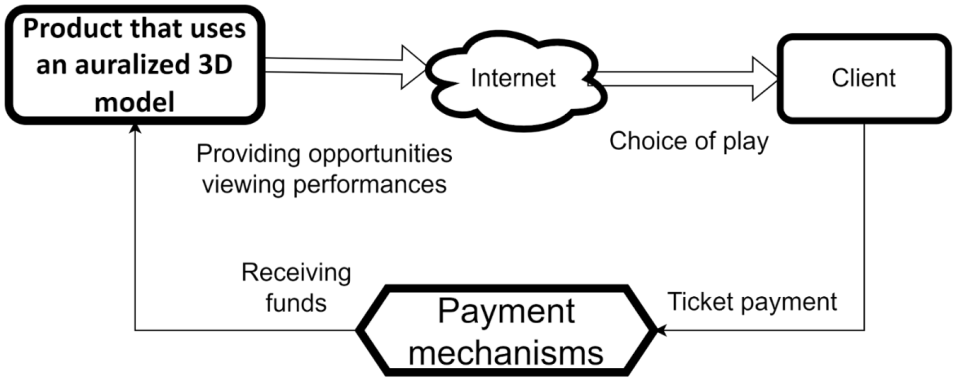


Fig. 1. E-commerce business process on the example of buying a ticket for a performance. Source [9] (modified by the authors)

In *block 5,* product revenues from an auralised 3D model can be derived from government allocations, for example, for the development of cultural heritage. At the same time, an auralised 3D model as a product can earn money on its own through various monetisation routes [16-19]:

1. advertising in the product;
2. additional options: in-app purchases, subscriptions, payment for downloads;
3. creating a marketplace.

In the case of the *first* monetisation path, it is advisable to use one of the most common business processes - "Affiliations" (Fig. 2), which demonstrates the interaction between three entities: listeners, sellers and the affiliate network (Google Adwords, Facebook Ads) [20].

Data about user behaviour and habits may be collected and provided to advertisers for in-app advertising in the following forms:

- Banner advertising [21-23], which is placed in the form of banner posters when watching a performance;
- Native advertising, in which the user hovers over an object and has the opportunity to follow an advertising link. Examples for a product with an auralised 3D model include: the clothes of other viewers, furniture on stage, musical instruments, etc;
- An intermediate full-screen web page with ads displayed before or after the expected page;
- Video advertising in a product with an auralised 3D model that can be displayed at all possible viewing levels.

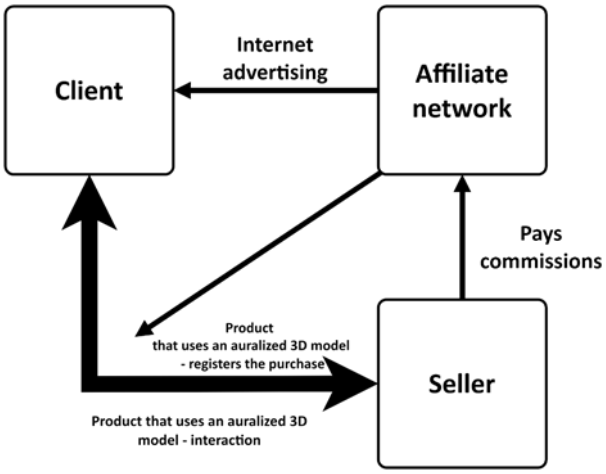


Fig. 2. Business process "Affiliation". Source [9] (modified by the authors)

In the case of the *second* monetisation path (purchase of goods), it is advisable to consider the business process "Additional features" [24], which usually includes a fairly well-thought-out pricing strategy [25]. In a product that uses an auralised 3D model, the user can choose additional features (Fig. 3). When forming the value proposition of a ticket to a performance, it is necessary to determine which additional features can bring maximum revenue.

Using the *third* monetisation path, new markets or marketplaces can be created to connect listeners with performers of musical works with an auralised 3D model. In addition, there is an opportunity to develop a marketplace that can be joined by other theatres around the world, providing relevant services and paying interest for the use of the auralised 3D model.

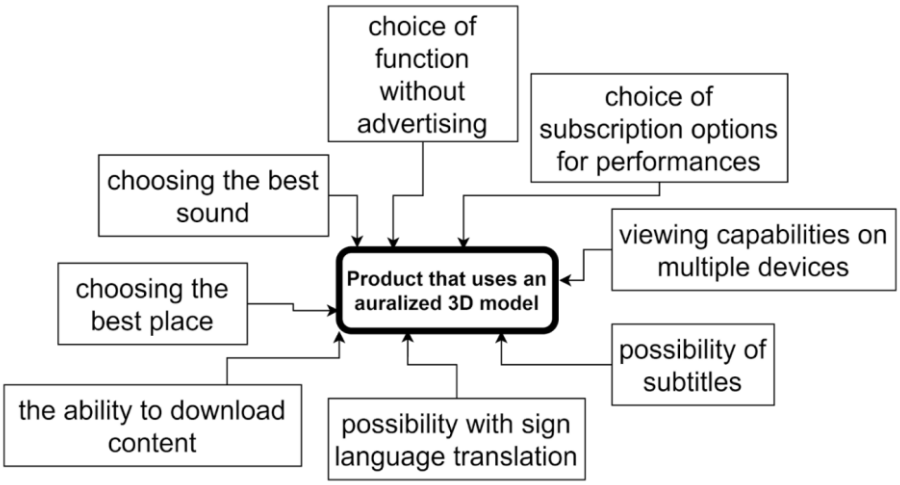


Fig. 3. Business process "Additional opportunities". Source [9] (modified by the authors)

Business model 2 "Barter"

One of the most attractive on the Internet is the Pay with a Tweet system based on the Barter business model, which uses social media platforms to promote products and services [9].

The application of this business model to a product using an auralised 3D model makes it possible to create a potential base of classical music listeners and, accordingly, to form a feedback base. Moreover, it is possible to test future products with an auralised 3D model. Figure 6 shows an example of how the Barter business process is used to interact with classical music listeners and social media communities with a product that uses an auralised 3D model.

Business model 3 "Joint ownership"

Under the Shared Ownership business model, investors do not purchase the entire asset, but only a portion of it. At the same time, the owners of the product using the auralised 3D model manage the assets (capital). Applying this business model to a product that uses an auralised 3D model will create an investment environment. Figure 4 shows an example of how potential investors interact with the owner of a product that uses an auralised 3D model within the Shared Ownership business process.

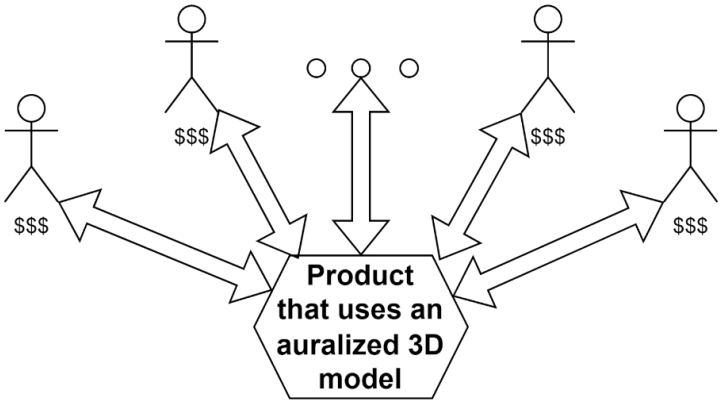


Fig. 4. Business process "Joint ownership". Source [9] (modified by the authors)

Simulation Model of Business Processes for Auralisation Product Using System Dynamics

Based on the business processes described above and system dynamics methods [26], and taking into account previous experimental studies, the structure of the simulation model was developed (Fig. 5). It will allow us to predict the revenue stream from the production of an auralised 3D model in the form of a VR application for theatre. The structure shows that each business process is a separate stream: Affiliation, Barter, E-commerce, Joint_ownership, MONEY.

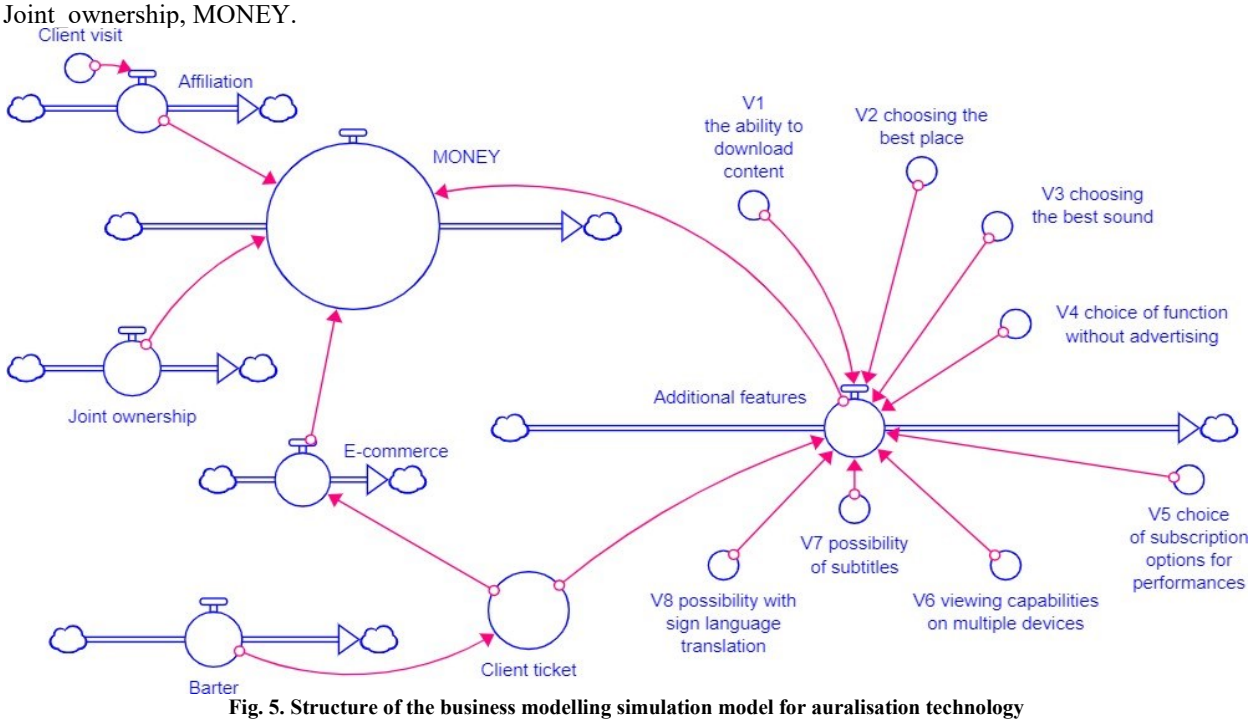


Fig. 5. Structure of the business modelling simulation model for auralisation technology

The Affiliation flow is dependent on the Client visit parameter. The Barter flow, in the definition of a money flow, has only an impact on the number of people who have purchased tickets (Client ticket). E-commerce is affected only by the Client ticket.

The Joint_ownership stream has the most parameters to influence, because many different possibilities can be provided when listening to a performance in a theatre based on auralisation technology. For example, it can be: V1 the ability to download content; V2 choosing the best location; V3 choosing the best sound; V4 choosing a function without advertising; V5 choosing subscription options for performances; V6 viewing capabilities on multiple devices; V7 possibility of subtitles; V8 possibility with sign language translation.

Conclusions.

The authors developed business models for the product using the auralised 3D model. Business processes are also presented within the business models, in particular: Affiliation, Barter, E-commerce, Joint_ownership.

Based on the business processes the structure of the simulation model was proposed. The outcomes can be extended and adapted to other products.

In further, authors are going to run the scenario modelling using system dynamics approaches as well as in-depth analysis of marketing research employing a number of demonstrative experiences.

References

1. Results / AURA – Auralisation of acoustic heritage sites using Augmented and Virtual Reality. AURA – Auralisation of acoustic heritage sites using Augmented and Virtual Reality. https://aura-project.eu/en/results/
2. Baden-Fuller, C., & Haefliger, S. (2013). Business models and technological innovation. Long range planning, Elsevier, 46(6), 419-426
3. Osterwalder, A., & Pigneur, Y. (2010). *Business model generation: a handbook for visionaries, game changers, and challengers* (Vol. 1). John Wiley & Sons.
4. Demil, B., & Lecocq, X. (2010). Business model evolution: in search of dynamic consistency. Long range planning, Elsevier, 43(2-3), 227-246.
5. R. Bärenfänger and B. Otto, "Proposing a Capability Perspective on Digital Business Models," 2015 IEEE 17th Conference on Business Informatics, 2015, pp. 17-25, doi: 10.1109/CBI.2015.18.
6. Pourbafrani, M., van Zelst, S.J., van der Aalst, W.M.P. (2019). Scenario-Based Prediction of Business Processes Using System Dynamics. In: Panetto, H., Debruyne, C., Hepp, M., Lewis, D., Ardagna, C., Meersman, R. (eds) On the Move to Meaningful Internet Systems: OTM 2019 Conferences. OTM 2019. Lecture Notes in Computer Science(), vol 11877. Springer, Cham. https://doi.org/10.1007/978-3-030-33246-4_27
7. De Backer, M., Snoeck, M., Monsieur, G., Lemahieu, W., & Dedene, G. (2009). A scenario-based verification technique to assess the compatibility of collaborative business processes. Data & knowledge engineering, 68(6), 531-551.
8. Bocken, N., Short, S., Rana, P., & Evans, S. 2014. A literature and practice review to develop sustainable business model archetype. Journal of Cleaner Production, 65: 42-56.
9. Gassmann, O., Frankenberger, K., & Csik, M. (2014). *The business model navigator: 55 models that will revolutionise your business*. Pearson UK.
10. Remane, G., Hanelt, A., Tesch, J., & Kolbe, L. 2017. The Business Model Pattern Database A Tool for Systematic Business Model Innovation. International Journal of Innovation Management, 21(1): Art. No. 1750004.
11. Al-Fedaghi, S., Makdessi, M. (2020). Modeling Business Process and Events. In: Silhavy, R. (eds) Intelligent Algorithms in Software Engineering. CSOC 2020. Advances in Intelligent Systems and Computing, vol 1224. Springer, Cham. https://doi.org/10.1007/978-3-030-51965-0_32
12. Blank, S. (2013), Why the Lean Start-Up Changes Everything, Harvard Business Review, 91(5), 63-72.
13. Müller, R.M., Thoring, K. (2012), Design Thinking vs. Lean Startup: A Comparison of Two User-Driven Innovation Strategies, Leading Through Design, Proceedings of the DMI 2012 International Research Conference, August 8-9, Boston, MA.
14. Lima, Marcos & Baudier, Patricia. (2017). Business Model Canvas Acceptance among French Entrepreneurship Students: Principles for Enhancing Innovation Artefacts in Business Education. Journal of Innovation Economics. 23. 159. 10.3917/jie.pr1.0008.
15. Qin, Z., & Qin, Z. (2009). Introduction to E-commerce (Vol. 2009). Heidelberg: Springer.
16. Crespo, M., Pinto-Martinho, A., Foà, C., Paisana, M., & Pais, P. C. (2020). Business models of journalistic startups in Portugal: an analysis of product innovation, dissemination and monetization in media enterprises. Business models of journalistic startups in Portugal: an analysis of product innovation, dissemination and monetization in media enterprises, (2), 261-296.
17. Portilla, I., & Medina, M. (2016). Monetization strategies and audience data for online video. The case of Atresmedia. Quaderns del CAC, 19(42), 27-36.
18. Li, YM., Fang, Y., Jin, BH. (2012). Analyzing Monetization Models for the Digital Content Services: Channel Ownership and Royalty Contracts. In: Shaw, M.J., Zhang, D., Yue, W.T. (eds) E-Life: Web-Enabled Convergence of Commerce, Work, and Social Life. WEB 2011. Lecture Notes in Business Information Processing, vol 108. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-29873-8_18
19. Bhattacharya, S., Scott, E. & Arthur, M. The phoenix rises from the ashes: Advertising and content monetization in a digital world. J Digit Asset Manag 2, 269–278 (2006). https://doi.org/10.1057/palgrave.dam.3650046
20. Michalis Vafopoulos (2012), "The Web Economy: Goods, Users, Models, and Policies", Foundations and Trends® in Web Science: Vol. 3: No. 1–2, pp 1-136. http://dx.doi.org/10.1561/1800000015
21. HOW DO FREE APPS MAKE MONEY | 8 Best monetization models for App monetization on Apple app store and Google play store. (б. д.). Best Web & Mobile App Developers Company | The APP Solutions. https://theappsolutions.com/blog/marketing/free-app-money/
22. Reklama uniwersytetu przez portal edukacyjny. https://kudapostupat.ua/reklama-banerowa
23. Robertson, A. (2021, 16 червня). Facebook will start putting ads in Oculus Quest apps. The Verge. https://www.theverge.com/2021/6/16/22535511/facebook-ads-oculus-quest-vr-app
24. T. A. -D. Hoang and T. B. Nguyen, "State of the Art and Emerging Rule-driven Perspectives towards Service-based Business Process Interoperability," 2009 IEEE-RIVF International Conference on Computing and Communication Technologies, 2009, pp. 1-4, doi: 10.1109/RIVF.2009.5174659
25. Harika, A., Sunil Kumar, M., Anantha Natarajan, V., Kallam, S. (2021). Business Process Reengineering: Issues and Challenges. In: Goyal, D., Chaturvedi, P., Nagar, A.K., Purohit, S. (eds) Proceedings of Second International Conference on Smart Energy and Communication. Algorithms for Intelligent Systems. Springer, Singapore. https://doi.org/10.1007/978-981-15-6707-0_35
26. Ford, D. N., & Lyneis, J. M. (2020). System dynamics applied to project management: a survey, assessment, and directions for future research. System Dynamics: Theory and Applications, Springer, 285-314.

Khrystyna Lipianina-Honcharenko **Христина Ліп'яніна-Гончаренко**	Associate professor, Ph.D. in information technologies Department for Information Computer Systems and Control, West Ukrainian National University https://orcid.org/0000-0002-2441-6292 e-mail: xrustya.com@gmail.com	Кандидат технічних наук, доцент кафедри інформаційно-обчислювальних систем і управління, Західноукраїнський національний університет
Anatoliy Sachenko **Анатолій Саченко**	Professor, Department for Information Computer Systems and Control, West Ukrainian National University ORCID: 0000-0002-0907-3682 e-mail: as@wunu.edu.ua	Доктор технічних наук, професор кафедри інформаційно-обчислювальних систем і управління, Західноукраїнський національний університет

Vitaliy Kulyk **Віталій Кулик**	Director of Lviv Tourism Development Center - Municipal Enterprise e-mail: vitkulyk@gmail.com	Директор ЛКП "Центр розвитку туризму м. Львова"
Ruslan Savchyshyn **Руслан Савчишин**	President MagneticOne e-mail: ruslan@magneticone.com	Президент MagneticOne
Oleksandra Provozin **Олександра Провозін**	Chief specialist of Lviv Tourism Development Center - Municipal Enterprise e-mail: oprovozin@gmail.com	Головний спеціаліст ЛКП "Центр розвитку туризму м. Львова"
Svitlana Shchur **Світлана Щур**	CFO MagneticOne e-mail: s.shchur@magneticone.com	Фінансовий директор MagneticOne
Liliia Kurpita **Лілія Курпіта**	Deputy Director of Lviv Tourism Development Center - Municipal Enterprise e-mail: liliiakurpita.work@gmail.com	Заступник директора ЛКП "Центр розвитку туризму м. Львова"

Full requirements for the design of the manuscript
Повні вимоги до оформлення рукопису
http://csitjournal.khmnu.edu.ua/

No editorial responsibility is required for the content of messages sub.
За зміст повідомлень редакція відповідальності не несе

To print 29.12.2022. Mind. Printing. Arch. 9,43. Obl.-vid. Arch. 9,04
Format 30x42 / 4, offset paper. Another risography.
Overlay 100, deputy. №

Підп. до друку 29.12.2022. Ум. друк. арк. 9,43. Обл.-вид. арк. 9.04
Формат 30х42/4, папір офсетний. Друк різографією.
Наклад 100, зам. №

Replication is made from the original layout, made edited
by the magazine "Computer Systems and Information Technology"

Тиражування здійснено з оригінал-макету, виготовленого
редакцією журналу “Комп’ютерні системи та інформаційні технології”

Editorial and publishing center of Khmelnytskyi national university
29016, Khmelnytskyi, street Institutska, 7/1, tel. (0382) 72-83-63

Редакційно-видавничий центр Хмельницького національного університету
29016, м. Хмельницький, вул. Інститутська, 7/1, тел. (0382) 72-83-63