

# 3rd Workshop for Young Scientists in Computer Science & Software Engineering

Arnold E. Kiv<sup>a</sup>, Serhiy O. Semerikov<sup>b,c,d</sup>, Vladimir N. Soloviev<sup>b,e</sup> and Andrii M. Striuk<sup>c</sup>

<sup>a</sup>Ben-Gurion University of the Negev, P.O.B. 653, Beer Sheva, 8410501, Israel

<sup>a</sup>Kryvyi Rih State Pedagogical University, 54 Gagarin Ave., Kryvyi Rih, 50086, Ukraine

<sup>b</sup>Kryvyi Rih National University, 11 Vitalii Matusevych Str., Kryvyi Rih, 50027, Ukraine

<sup>c</sup>Institute of Information Technologies and Learning Tools of the NAES of Ukraine, 9 M. Berlynskoho Str., Kyiv, 04060, Ukraine

<sup>e</sup>Bohdan Khmelnytsky National University of Cherkasy, 81 Shevchenko Blvd., Cherkasy, 18031, Ukraine

## Abstract

This is an introductory text to a collection of selected papers from the 3rd Workshop for Young Scientists in Computer Science & Software Engineering (CS&SE@SW 2020), which was held in Kryvyi Rih, Ukraine, on the November 27, 2020. It consists of short summaries of selected papers and some observations about the event and its future.

## Keywords

computer science, software engineering, young scientists

## 1. CS&SE@SW 2020: At a glance

Workshop for Young Scientists in Computer Science & Software Engineering (CS&SE@SW) is a peer-reviewed workshop focusing on research advances, applications of information technologies.

CS&SE@SW topics of interest since 2018 [1, 2] are:

- Computer Science (CS):
  - Theoretical computer science
    - \* Data structures and algorithms
    - \* Theory of computation
    - \* Information and coding theory

*CS&SE@SW 2020: 3rd Workshop for Young Scientists in Computer Science & Software Engineering, November 27, 2020, Kryvyi Rih, Ukraine*

✉ [kiv.arnold20@gmail.com](mailto:kiv.arnold20@gmail.com) (A.E. Kiv); [semerikov@gmail.com](mailto:semerikov@gmail.com) (S.O. Semerikov); [vnsoloviev2016@gmail.com](mailto:vnsoloviev2016@gmail.com) (V.N. Soloviev); [andrey.n.stryuk@gmail.com](mailto:andrey.n.stryuk@gmail.com) (A.M. Striuk)

🌐 <https://kdpu.edu.ua/semerikov> (S.O. Semerikov); <https://kdpu.edu.ua/personal/vmsoloviev.html> (V.N. Soloviev); <http://mpz.knu.edu.ua/pro-kafedru/vikladachi/224-andrii-striuk> (A.M. Striuk)

🆔 0000-0002-0991-2343 (A.E. Kiv); 0000-0003-0789-0272 (S.O. Semerikov); 0000-0002-4945-202X (V.N. Soloviev); 0000-0001-9240-1976 (A.M. Striuk)

© 2020 Copyright for this paper by its authors.  
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).



CEUR Workshop Proceedings (CEUR-WS.org)

- \* Programming language theory
- \* Formal methods
- Computer systems
  - \* Computer architecture and computer engineering
  - \* Computer performance analysis
  - \* Concurrent, parallel and distributed systems
  - \* Computer networks
  - \* Formal methods
  - \* Databases
- Computer applications
  - \* Computer graphics and visualization
  - \* Human-computer interaction
  - \* Scientific computing
  - \* Artificial intelligence
- Software Engineering (SE):
  - Software requirements
  - Software design
  - Software construction
  - Software testing
  - Software maintenance
  - Software configuration management
  - Software engineering management
  - Software development process
  - Software engineering models and methods
  - Software quality
  - Software engineering professional practice
  - Software engineering economics
  - Computing foundations
  - Mathematical foundations
  - Engineering foundations

This volume represents the proceedings of the 3rd Workshop for Young Scientists in Computer Science & Software Engineering (CS&SE@SW 2020), held in Kryvyi Rih, Ukraine, on November 27, 2020. It comprises 9 contributed papers that were carefully peer-reviewed and selected from 15 submissions. Each submission was reviewed by at least 2, and on the average 2.7, program committee members. The accepted papers present the state-of-the-art overview of successful cases and provides guidelines for future research.



## **2. CS&SE@SW 2020 Program Committee**

### **2.1. Core Program Committee**

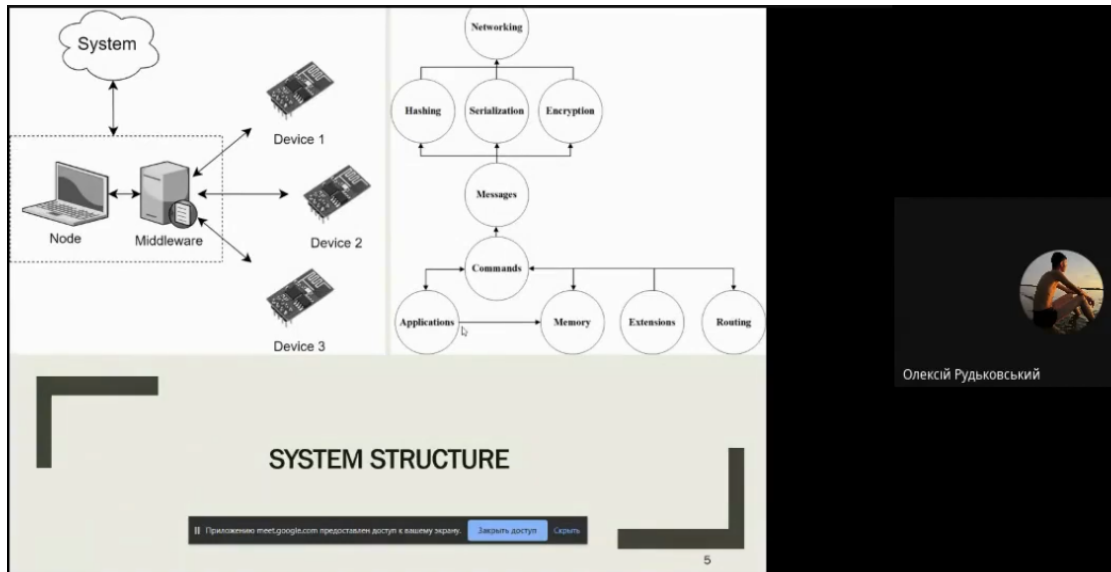
- *Arnold Kiv*, Ben-Gurion University of the Negev, Israel
- *Vasyl Oleksiuk*, Ternopil Volodymyr Hnatiuk National Pedagogical University, Ukraine
- *Viacheslav Osadchyi*, Bogdan Khmelnytsky Melitopol State Pedagogical University, Ukraine
- *Antonii Rzheuskyi*, Lviv Polytechnic National University, Ukraine
- *Serhiy Semerikov*, Kryvyi Rih State Pedagogical University, Ukraine
- *Vladimir N. Soloviev*, Kryvyi Rih State Pedagogical University, Ukraine

### **2.2. Additional reviewers**

- *Pavlo Hryhoruk*, Khmelnytskyi National University, Ukraine
- *Oleksii Ignatenko*, Institute of Software Systems, Ukraine
- *Oleksandr Kolgatin*, Simon Kuznets Kharkiv National University of Economics, Ukraine
- *Andrey Kupin*, Kryvyi Rih National University, Ukraine
- *Oleg Pursky*, Kyiv National University of Trade and Economics, Ukraine
- *Andrii Striuk*, Kryvyi Rih National University, Ukraine
- *Nataliia Veretennikova*, Lviv Polytechnic National University, Ukraine

## **3. CS&SE@SW 2020 Article overview**

Oleksii R. Rudkovskyi (figure 1) and Galina G. Kirichek in their article [3] implemented a method of organizing a distributed network to launch and support the work of applications. In the process of building the network model, different algorithms were used, at the same time



**Figure 1:** Presentation of paper [3]

encryption algorithms are used (Elliptic curve Diffie–Hellman and Advanced Encryption Standard), Secure Hash Algorithm 1 algorithms for hashing, onion routing algorithm and other. Transmission Control Protocol transport protocol is used to send data between nodes of the network. Methods and application for creating and launching other applications have been implemented in the Java language. Linux containers and Docker are used to isolate different applications running on the network. In this way, different applications written using different programming languages can run without affecting other running applications and client's operation system. Application's data is stored on the network distributed with and without encryption.

Vladimir N. Soloviev, Andrii O. Bielinskyi (figure 2) and Natalia A. Kharadzjan in the article [4] demonstrate the possibility of constructing indicators of critical and crash phenomena on the example of Bitcoin market crashes for further demonstration of their efficiency on the crash that is related to the coronavirus pandemic. For this purpose, the methods of the theory of complex systems have been used. Since the theory of complex systems has quite an extensive toolkit for exploring the nonlinear complex system, authors take a look at the application of the concept of entropy in finance and use this concept to construct 6 effective entropy measures: Shannon entropy, Approximate entropy, Permutation entropy, and 3 Recurrence based entropies. Authors provide computational results that prove that these indicators could have been used to identify the beginning of the crash and predict the future course of events associated with the current pandemic.

Reducing costs is an important part in today's business. Therefore manufacturers try to reduce unnecessary work processes and storage costs. Machine maintenance is a big, complex, regular process. In addition, the spare parts required for this must be kept in stock until a machine fails. In order to avoid a production breakdown in the event of an unexpected failure,

## Recurrence entropy (Microstates) (Empirical results)

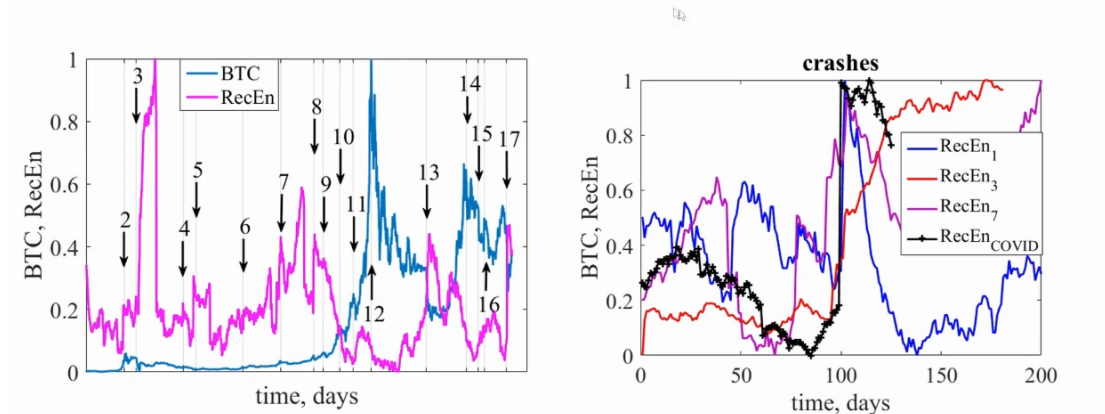


Figure 2: Presentation of paper [4]

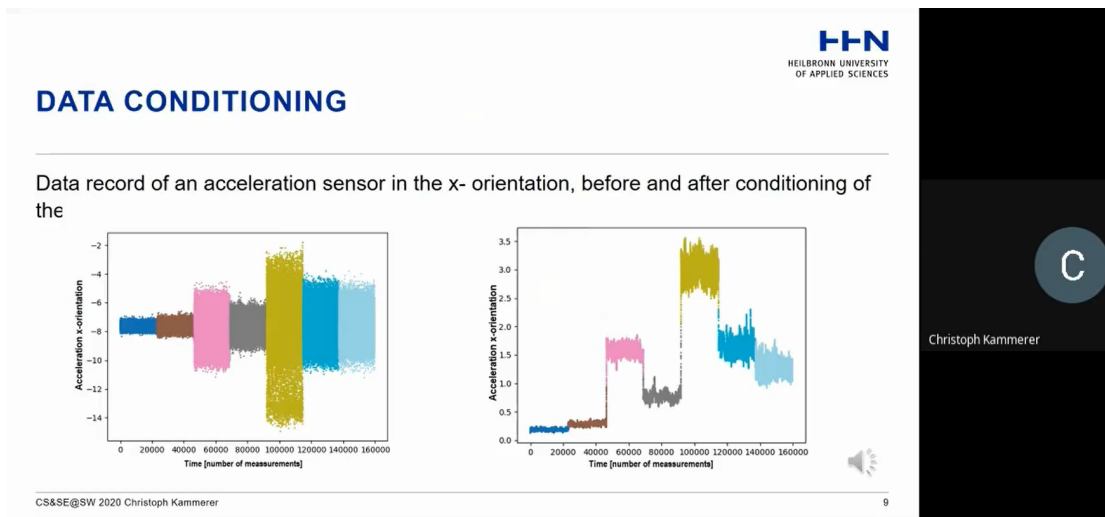


Figure 3: Presentation of paper [5]

more and more manufacturers rely on predictive maintenance for their machines. This enables more precise planning of necessary maintenance and repair work, as well as a precise ordering of the spare parts required for this. A large amount of past as well as current information is required to create such a predictive forecast about machines. With the classification of motors based on vibration, the paper [5] of Christoph Kammerer (figure 3), Micha Küstner, Michael Gaust, Pascal Starke, Roman Radtke and Alexander Jesser deals with the implementation of predictive maintenance for thermal systems. There is an overview of suitable sensors and data processing methods, as well as various classification algorithms. In the end, the best sensor-algorithm combinations are shown.

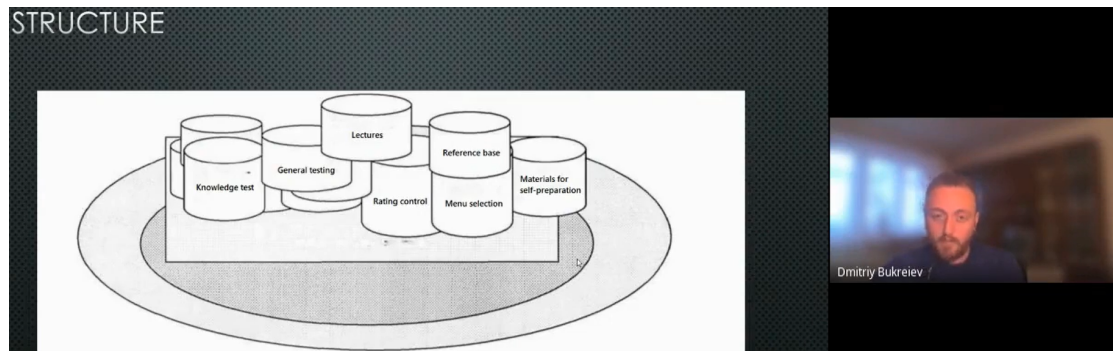


Figure 4: Presentation of paper [6]

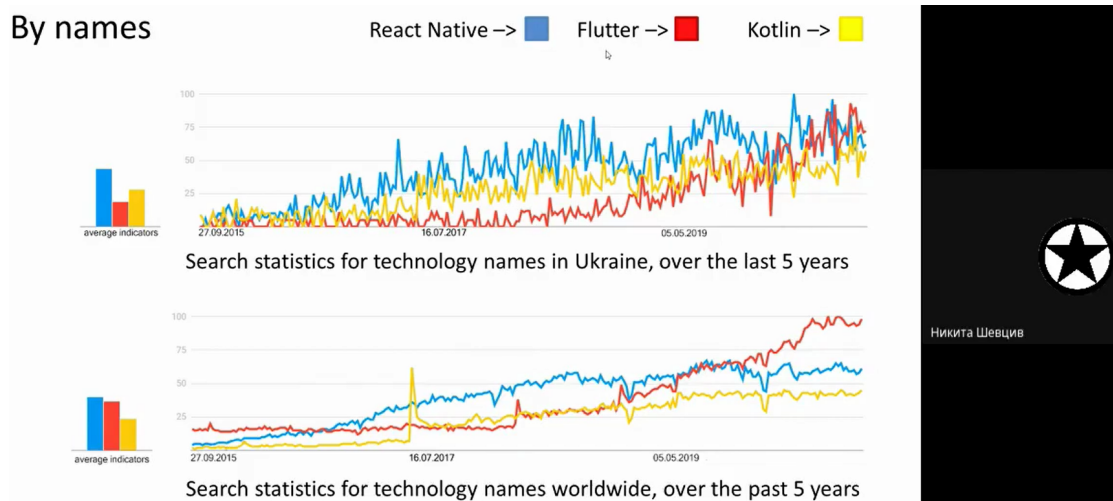


Figure 5: Presentation of paper [7]

The paper [6] of Dmitriy Bukreiev (figure 4), Pavlo Chorny, Evgeniy Kupchak and Andrey Sender reveals the problematic features of developing an automated system for teaching students. The authors emphasize the need to study the specialized needs of the system before the development process and conduct an in-depth analysis of the needs of the modern educational process in order to identify the needs of the future product. In this paper authors develop a model of knowledge representation in the system and reveal the features of the mathematical apparatus for calculating and automating the processes of obtaining and evaluating students knowledge.

The paper [7] of Nikita A. Shevtsiv (figure 5) and Andrii M. Striuk analyzes the advantages and disadvantages of cross-platform and native mobile application development. The conditions are highlighted in which native and cross-platform development reveal their advantages. These conditions include the project size, work comfort, popularity, relevance. It was concluded that a beginner developer should start learning from native development, and then try cross- platform.

## Tracking SDK related to WebGL trios

Babylon AR = Babylon.js + OpenCV: QR-codes at WebXR devices

JSARToolKit = ARToolKit + Three.js: bar codes, NFT markers

AR.js = AFrame + JSARToolKit: all in one

```
<a-scene embedded arjs>  
  <a-marker preset="hiro">  
    <a-box></a-box>  
  </a-marker>  
  <a-entity camera></a-entity>  
</a-scene>
```



**Figure 6:** Presentation of paper [8]

Web augmented reality (WebAR) development tools aimed at improving the visual aspects of learning are far from being visual and available themselves. This causing problems of selecting and testing WebAR development tools for CS undergraduates mastering in web-design basics. The article [8] of Dmytro S. Shepiliev, Yevhenii O. Modlo, Yuliia V. Yechkalo, Viktoriia V. Tkachuk, Mykhailo M. Mintii, Iryna S. Mintii, Oksana M. Markova, Tetiana V. Selivanova, Olena M. Drashko, Olga O. Kalinichenko, Tetiana A. Vakaliuk, Viacheslav V. Osadchyi and Serhiy O. Semerikov (figure 6) is aimed at conducting comparative analysis of WebAR tools to select those appropriated for beginners.

Machine learning is now widely used almost everywhere, primarily for forecasting. The main idea of the article [9] of Pavlo V. Zahorodko, Yevhenii O. Modlo, Olga O. Kalinichenko, Tetiana V. Selivanova and Serhiy O. Semerikov (figure 7) is to identify the possibility of achieving a quantum advantage when solving machine learning problems on a quantum computer.

The paper [10] of Bohdan V. Hrebeniuk (figure 8) and Olena H. Rybalchenko analyzes the existing platforms for conducting programming contests. Possible approaches are analyzed for creating isolated environments and running participants' solutions, advantages and disadvantages of both approaches are highlighted. Requirements for the user interface are defined that must provide quick and convenient work in the system; the system was planned and developed. It was concluded that designed system has a potential for conducting contests and further development.

The paper [11] of Mykola V. Klymenko (figure 9) and Andrii M. Striuk considers the typical technical features of GPS-tracking systems and their development, as well as an analysis of existing solutions to the problem. Mathematical models for the operation of hardware and software of this complex have been created. An adaptive user interface has been developed that allows you to use this complex from a smartphone or personal computer. Methods for

Criteria	Indicators
It arises from scientific curiosity about the foundation, the nature or the limits of a scientific discipline	Quantum computation has broadened the fundamental limits of computer science and software engineering
The ability to create new engineering solutions	The physical infrastructure is constantly evolving, each solution is new
Technological continuity	The existence of high level languages and development techniques that can be used by computer scientists and software engineers with only the same style of training they receive today (so, no need to teach the fundamentals of quantum mechanics to all)
Research community support	Support for all interested in new computing paradigms and new levels of computing power
International character of research	This is a new fundamental area of software engineering
It is generally comprehensible, and captures the imagination of the general public, as well as the esteem of scientists in other disciplines	It is not generally understood, but is known for its worldwide interpretation
The problem has a long-standing statement, but has not yet been resolved	Formulated by Richard Feynman in the late 1970s.
It promises to go beyond what is initially possible, and requires development of understanding, techniques and tools unknown at the start of the project	Problems exist on every level, from developing a whole new conceptual paradigm, to building intellectual and simulation tools
It calls for planned co-operation among identified research teams and communities	Research is needed in a number of areas (languages, algorithms, tools, simulation, visualisation, etc.)
It encourages and benefits from competition among individuals and teams, with clear criteria on who is winning, or who has won	There need not be a single "winner", diversity of solutions should be encouraged, as in classical software engineering, to be applicable to a range of application domains
It decomposes into identified intermediate research goals, whose achievement brings scientific or economic benefit, even if the project as a whole fails	There are several components of the problem that can be explored in parallel
It will lead to radical paradigm shift	Quantum computing is a radical paradigm shift

Figure 7: Presentation of paper [9]

## Docker usage

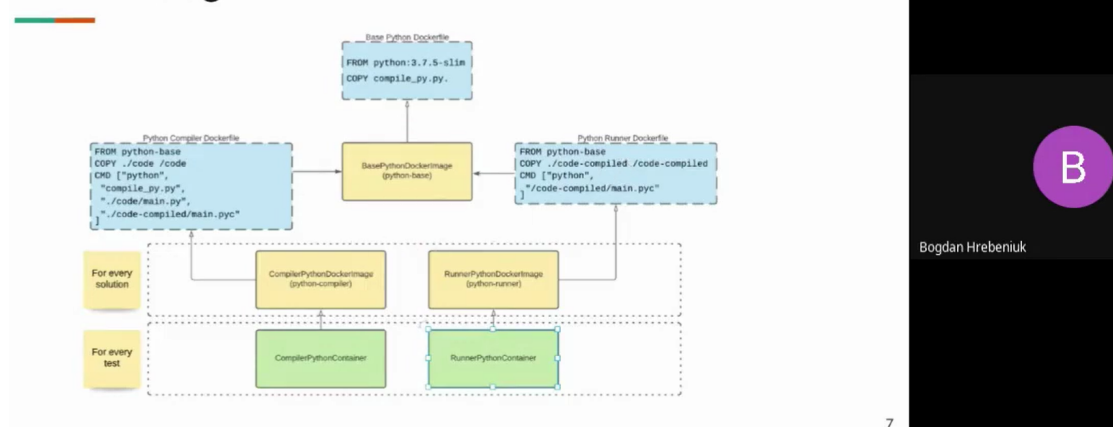


Figure 8: Presentation of paper [10]

displaying the distance traveled by a moving object on an electronic map have been developed. Atmega162-16PU microcontroller software for GSM module and GPS receiver control has been developed. A method of data transfer from a GPS tracker to a web server has been developed. Two valid experimental samples of GPS-trackers were made and tested in uncertain conditions. The GPS-tracking software and hardware can be used to monitor the movement of moving objects that are within the coverage of GSM cellular networks.



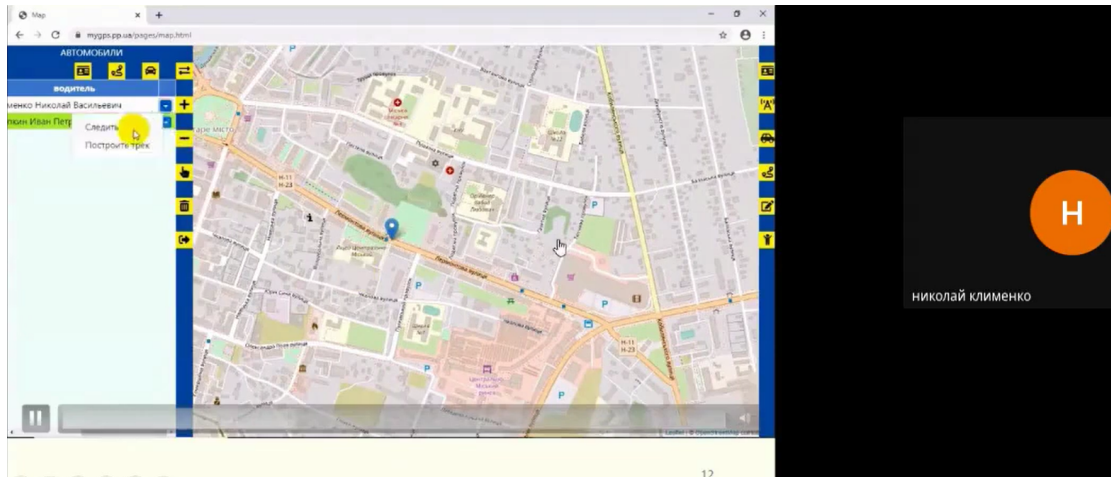


Figure 9: Presentation of paper [11]

## 4. CS&SE@SW 2020: Conclusion and outlook

The vision of the CS&SE@SW 2020 is provides an expert environment for young researchers, who are at the beginning of their career, to present and discuss the most recent of ideas and early results of research projects. Young researchers, who will join us to take part in discussions and/or present their papers, will be offered an opportunity to exchange and discuss their research ideas with their peers, supervisors, and senior scientists working in the fields that are within the scope of CS&SE@SW.

The third instalment of CS&SE@SW was organised by Kryvyi Rih National University, Ukraine (with support of the rector Mykola I. Stupnik) in collaboration with Kryvyi Rih State Pedagogical University, Ukraine (with support of the rector Yaroslav V. Shramko), Institute of Information Technologies and Learning Tools of the NAES of Ukraine (with support of the director Valeriy Yu. Bykov) and Ben-Gurion University of the Negev, Israel (with support of the rector Chaim J. Hames).

We are thankful to all the authors who submitted papers and the delegates for their participation and their interest in CS&SE@SW as a platform to share their ideas and innovation. Also, we are also thankful to all the program committee members for providing continuous guidance and efforts taken by peer reviewers contributed to improve the quality of papers provided constructive critical comments, improvements and corrections to the authors are gratefully appreciated for their contribution to the success of the workshop.

We hope you enjoy this workshop and meet again in more friendly, hilarious, and happiness of further CS&SE@SW 2021 at Kryvyi Rih, Ukraine on November 26, 2021.

## References

- [1] A. Kiv, S. Semerikov, V. Soloviev, A. Striuk, First student workshop on computer science & software engineering, volume 2292, CEUR-WS, 2018, pp. 1–10. 1st Student Workshop

- on Computer Science and Software Engineering, CS and SE@SW 2018 ; Conference Date: 30 November 2018.
- [2] A. Kiv, S. Semerikov, V. Soloviev, A. Striuk, Second student workshop on computer science & software engineering, volume 2546, CEUR-WS, 2019, pp. 1–20. 2nd Student Workshop on Computer Science and Software Engineering, CS and SE@SW 2019 ; Conference Date: 29 November 2019.
  - [3] O. R. Rudkovskiy, G. G. Kirichek, Interaction support system of network applications, CEUR-WS, 2020, pp. 11–23. 3rd Workshop for Young Scientists in Computer Science & Software Engineering, CS&SE@SW 2020 ; Conference Date: 27 November 2020.
  - [4] V. N. Soloviev, A. O. Bielinskyi, N. A. Kharadzjan, Coverage of the coronavirus pandemic through entropy measures, CEUR-WS, 2020, pp. 24–42. 3rd Workshop for Young Scientists in Computer Science & Software Engineering, CS&SE@SW 2020 ; Conference Date: 27 November 2020.
  - [5] C. Kammerer, M. Küstner, M. Gaust, P. Starke, R. Radtke, A. Jesser, Classification of motor vibration with machine learning methods and simulating the vibration using statistical models, CEUR-WS, 2020, pp. 43–54. 3rd Workshop for Young Scientists in Computer Science & Software Engineering, CS&SE@SW 2020 ; Conference Date: 27 November 2020.
  - [6] D. Bukreiev, P. Chorny, E. Kupchak, A. Sender, Features of the development of an automated educational and control complex for checking the quality of students, CEUR-WS, 2020, pp. 55–74. 3rd Workshop for Young Scientists in Computer Science & Software Engineering, CS&SE@SW 2020 ; Conference Date: 27 November 2020.
  - [7] N. A. Shevtsiv, A. M. Striuk, Cross platform development vs native development, CEUR-WS, 2020, pp. 75–83. 3rd Workshop for Young Scientists in Computer Science & Software Engineering, CS&SE@SW 2020 ; Conference Date: 27 November 2020.
  - [8] D. S. Shepiliev, Y. O. Modlo, Y. V. Yechkalo, V. V. Tkachuk, M. M. Mintii, I. S. Mintii, O. M. Markova, T. V. Selivanova, O. M. Drashko, O. O. Kalinichenko, T. A. Vakaliuk, V. V. Osadchyi, S. O. Semerikov, WebAR development tools: An overview, CEUR-WS, 2020, pp. 84–93. 3rd Workshop for Young Scientists in Computer Science & Software Engineering, CS&SE@SW 2020 ; Conference Date: 27 November 2020.
  - [9] P. V. Zahorodko, Y. O. Modlo, O. O. Kalinichenko, T. V. Selivanova, S. O. Semerikov, Quantum enhanced machine learning: An overview, CEUR-WS, 2020, pp. 94–103. 3rd Workshop for Young Scientists in Computer Science & Software Engineering, CS&SE@SW 2020 ; Conference Date: 27 November 2020.
  - [10] B. V. Hrebeniuk, O. H. Rybalchenko, Development of an automated system for conducting, checking and evaluating programming competitions, CEUR-WS, 2020, pp. 104–114. 3rd Workshop for Young Scientists in Computer Science & Software Engineering, CS&SE@SW 2020 ; Conference Date: 27 November 2020.
  - [11] M. V. Klymenko, A. M. Striuk, Development of software and hardware complex of GPS-tracking, CEUR-WS, 2020, pp. 115–129. 3rd Workshop for Young Scientists in Computer Science & Software Engineering, CS&SE@SW 2020 ; Conference Date: 27 November 2020.

# Interaction support system of network applications

Oleksii R. Rudkovskiy<sup>a</sup>, Galina G. Kirichek<sup>a,b</sup>

<sup>a</sup>Donetsk National Technical University, 2 Shybankova Sq., Pokrovsk, 85300, Ukraine

<sup>b</sup>National University "Zaporizhzhia Polytechnic", 64 Zhukovsky St., Zaporizhia, 69063, Ukraine

## Abstract

In this article implemented a method of organizing a distributed network to launch and support the work of applications. In the process of building the network model, different algorithms were used, at the same time encryption algorithms are used (Elliptic curve Diffie–Hellman and Advanced Encryption Standard), Secure Hash Algorithm 1 algorithms for hashing, onion routing algorithm and other. Transmission Control Protocol transport protocol is used to send data between nodes of the network. Methods and application for creating and launching other applications have been implemented in the Java language. Linux containers and Docker are used to isolate different applications running on the network. In this way, different applications written using different programming languages can run without affecting other running applications and client's operation system. Application's data is stored on the network distributed with and without encryption.

## Keywords

Docker, distributed networks, encryption, Java, TCP

## 1. Introduction

In today's world it is impossible to imagine the operation of data transmission devices for various purposes without a network connection. It is even more difficult to imagine performing a large number of tasks without using the usual services, such as e-mail, cloud storage, various messengers, web applications, etc. In addition, the development of data transmission systems has reached a level where almost any object can be connected to the network. This constantly increases the amount of traffic and data that needs to be stored or processed. Therefore, with the development of information technology and data transmission methods, the issues of further development of networks and support of secure data transmission systems are becoming more and more urgent and acute [1].

One of the problems of the modern Internet is the rapid development of the Internet of Things. According to the latest forecasts for the next few years, the number of devices connected to the Internet will increase several times. With the new version of the internet protocol (IP) IPv6, the number of addresses compared to IPv4, increased at times, but in the near future, this volume of addresses may be exhausted, according to the rate of growth of the Internet

---

CS&SE@SW 2020: 3rd Workshop for Young Scientists in Computer Science & Software Engineering, November 27, 2020, Kryvyi Rih, Ukraine

✉ alexteen12@gmail.com (O.R. Rudkovskiy); kirgal08@gmail.com (G.G. Kirichek)

🌐 <https://orcid.org/0000-0002-5267-2525> (O.R. Rudkovskiy); <http://csn.zntu.edu.ua/galina-grigorivna-kirichek> (G.G. Kirichek)

🆔 0000-0002-5267-2525 (O.R. Rudkovskiy); 0000-0002-0405-7122 (G.G. Kirichek)

© 2020 Copyright for this paper by its authors.  
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).



CEUR Workshop Proceedings (CEUR-WS.org)

of things. The reasons for the acceleration of filling IPv6 also include legislation of individual countries or alliances, the development of astronautics and the colonization of other planets [2].

Currently, almost all applications run in the form of client-server services. These services allow you to exchange messages, transfer files, receive news and publish them. Services can interact with both users and other services. But now the services are poorly protected and, after attacks by attackers, may lose confidential information left by users. In addition to public information, attackers steal bank card data, private files, etc., so the protection of this data is a priority for companies. In addition, it is constantly necessary to address the availability and speed of information transfer [3].

Thus, it is relevant to solve the problems of increasing the speed of information processing, supporting the security of transmission, storing and processing data, as well as executing server applications as close as possible to the client in order to reduce the delays between executing requests and receiving responses (critical for the operation of some categories of IoT devices).

## **2. Objectives and solutions**

### **2.1. Objectives**

The purpose of this work is to research and implement software that can support communication of network devices and applications, by developing protocol of distributed system to improve work on services and their communication with devices. The object of this research is the process of software implementation that support distributed communication between devices and applications. The subject of research are the models, methods and software tools for organizing the interaction of devices and applications in a distributed environment.

Currently, familiar virtual or dedicated machines (VPS and VDS) are used as a server infrastructure for applications, but recently containers are gaining more and more popularity due to the ease of scaling, security and speed of work [4, 5]. At the same time, the Internet is actively developing decentralized and distributed system on which it is also possible to create applications. These applications use a distributed database (usually blockchain), which guarantees the security of data storage, and allow interaction with applications with different devices and programs that are capable of performing certain calculations to participate in the system itself. The problem is that traditional technologies are not secure enough, and new distributed systems do not provide the same capabilities for creating and interacting with programs that traditional ones provide. For example, the Ethereum network provides the ability to develop applications that are able to receive data from outside and trigger events that other programs can process, but these programs are extremely limited in functionality and are not able to interact with other devices on the Ethereum network or with devices from Internet and are not capable of performing any resource-intensive tasks [6, 7].

The research is aimed at finding a solution that combines two approaches to developing and launching applications, providing a high level of security in the form of a distributed network and providing many opportunities for developers as traditional servers provides.

As the platform of distributed network many different independent devices (personal computers, mobile phones, servers, etc.), which contain launched specific software, is used. On

launching the application and while it works, devices connecting to each other and always ready to participate in processing or transferring a data. Device can support multiple roles: be a binder (intermediate) node, store data, launch applications or be a client who requests some calculations. Special algorithms are used to define a role for each node.

Before launching client's application, network should detect involved, active devices, find needed for processing requests nodes, connect involved nodes into one network and, after that, should launch needed applications. At the same time, while working, devices additionally divided to next types: computer node (executes needed application, in network can be one or more nodes of this type); client node (node, which requests calculations or that can receive calculated data); router (intermediate device that connect other nodes and control process of calculation).

For different types of nodes authors recommends to use different types of packets. In this case, router (intermediate device) should support safety of data transmission without having access to them in the same time. Based on the fact, that the network is distributed, we consider in advance any node in the network to be compromised and unsafe. Therefore, considering this fact, for solving safety problem when transferring data in distributed network in work proposed to use special algorithms for transferring and encrypting data.

## 2.2. Solutions for building a system

Consider several popular programming languages and applications that are suitable for realization of this task.

Java - a cross-platform programming language used for desktop and mobile devices supports a set of existing libraries for cryptography and network. Python is also a cross-platform, interpreted language, but slower. C++ is a low-level language, productive and with a large number of ready-made libraries, but more difficult to write program code [8]. After the analysis, the Java programming language was selected as cross-platform, object-oriented, popular, simple and fast enough [9, 10]. It allows you to run an application on many operating systems (Windows, Linux, macOS, etc.).

In the analysis of existing methods and encryption algorithms are considered the most popular AES, RSA, 3DES, and ECDH. A high level of security is provided by the asynchronous RSA or ECDH algorithm. At the same time, ECDH has high security with a shorter key length, and the RSA algorithm works faster. Since both algorithms do not allow encrypting large amounts of data, to solve the problem, we use a common encryption key and the synchronous AES algorithm as the safest and fastest [11, 12].

When choosing an isolated environment, virtualization methods using virtual machines and Linux containers are considered [2]. Virtualization consumes a lot of resources, which is why the number of concurrent virtual areas is very limited. Containers give access to all the resources of a node, are faster, and support the simultaneous operation of dozens of isolated areas [13]. When conducting research, Docker was chosen as an isolated environment, using Linux containers, as the most productive and not limiting the work of programs [13].

TCP is used as a data transfer protocol. For testing the system, a virtual machine created in VirtualBox with Ubuntu installed was selected. As a platform for transferring data overlay network is used. In this case, data transmission functions are assigned to the lower levels, thus

abstracted from the network, operating systems and technologies. The same data is sent over the same algorithm regardless of the kind of network technology used by the client.

The next functionality that should be implemented in the network is the self-organization of nodes. As nodes are independent of each other, network should automatically detect role of each node. For this purpose we use indicators of a node: free disk space (taking into account the limitations of each individual client of the network), installed software, number of processors and core, RAM size.

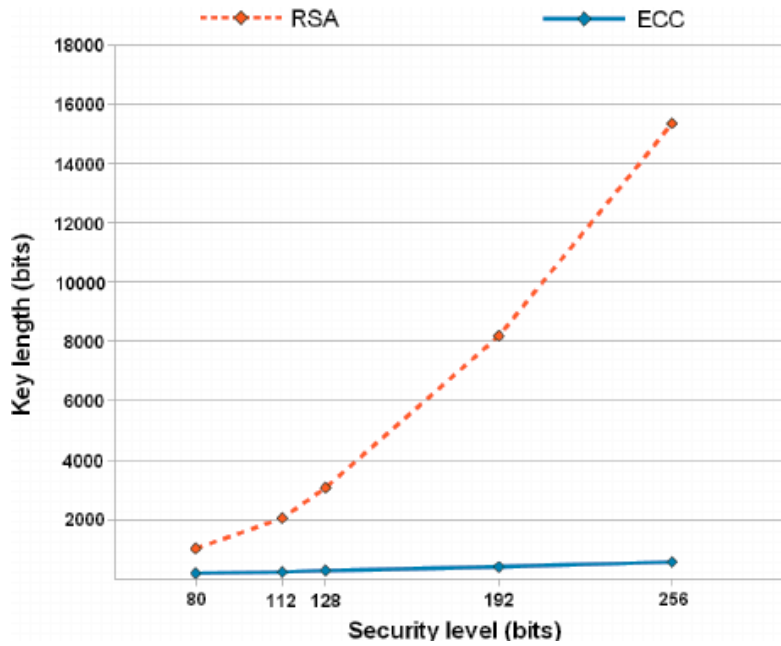
If storage is not full, by more than a certain percentage, node can receive files for storage, what he notify the nearest devices about. Otherwise, node work as distributor - he can send files on request, if they exists in local storage, but can't receive new files. In case if node should receive new file, but storage is full, node makes a decision to clear storage and delete unnecessary information. To do that, node asks the network for deleting a file. Only after receiving confirmation from some number of devices that file can be deleted, node deletes this file [14].

As different operating system and end nodes can launch different set of programs, a node store information about software that he can launch. Program, which should be executed, wherein store minimal set of needed and observed parameters. While launching a program, node make verification and, in case if his parameters are satisfy minimal requirements, a program starts. Otherwise, the request is redirected to the network until a node is found that can launch the application.

On a device at the same time can work many client's applications with different roles. Each application has a unique address within the network. With the constant growth of the number of devices connected to the global network, the use of existing routing methods, taking into account the peculiarities of the network being implemented, is impossible. Although IPv6 can satisfy a large number of users on the Internet, in the future, the process of combining and simultaneously using different networks (Ethernet and Bluetooth) may require an internal routing system. In this case, it is proposed to use the public key of the encryption algorithm as the node address in the implemented network. This approach allows both to securely encrypt data, since the encryption keys are not sent in the usual form, and to provide multiple addresses for one physical device. In addition, using this approach guarantees anonymity, since it is impossible to determine the user's location address, or any other information about the user.

Therefore, to generate a shared key, while device working in the network and encrypt a data, we use a pair keys (public and private) and public key of another client. Since the public key is an address of a device, and pair keys are stored only on one device and never passed on to anyone, this approach is most secure. After generating a shared key, the application encrypts the data sent with it. Second client, after generating a key using his own pair keys and address of the device who sent a data, will receive the same key [15].

In this case, if the sender and the recipient of the data are specified correctly, that is, the data is not sent from a foreign node, which is masked as the sender, the recipient can decrypt the data without any problems. This also applies to protect data from substitution or alteration, since in case of data change you also need to have access to the sender's key pair to generate the correct shared key and to encrypt the message itself. Moreover, if the data is distorted, the recipient will not be able to decrypt it, or will decrypt it with interference. Thus, the process of encryption and decryption makes the system itself sufficiently protected from interception



**Figure 1:** Comparison of RSA and ECDH algorithms

and data substitution.

For key pair and shared key generation cryptographic algorithm ECDH is used [16]. As you can see in figure 1, if compare to other algorithms, for example with RSA, the ECDH algorithm has a higher level of security with the same key size.

Although the ECDH algorithm is safe, it nevertheless does not allow encrypting large amounts of data since the maximum size of input data is limited by the key length. In addition, the creation of an encrypted message takes a long time, compared to other algorithms. To solve this problem, AES encryption algorithm and a shared key for the two devices are usually used. Algorithm AES can be considered cryptographically secure, and shared key generation without sharing parts of the keys through network makes it much secure and faster, depriving it of its main drawback - transmission of the key over the network [17, 18].

### 3. Stages of System Implementation

The nodes themselves can act as routers for devices that are unable to generate keys and encrypt data quickly enough. For example, microcontrollers cannot do all needed calculations fast enough, so as not to slow down its own functions.

#### 3.1. Process of generation key pair for devices

To solve this issue node, to which a similar device is connected to directly, generates new keys that are associated with this device. Thus, devices get their own keys and become indistinguishable for other clients. Algorithm of generating these keys illustrated in figure 2.

Using these algorithms and methods allows using of the implemented network with almost any end device, regardless of its own performance [19]. The used methods make it possible to safely transfer data in a given network, while the computation of data, at almost any end node of the system, minimizes the delay in data transmission between the client and the node that processes the data itself, since such a node can be a neighboring device located in close proximity to the client.

In addition, these methods and algorithms guarantee the anonymity of clients and adapted to work in the notoriously compromised environment.

### 3.2. Description of client software modeling process

During the research and before implementation, the client software was modeled and a diagram of the required classes was obtained, which is shown in figure 3.

For proper operation of the client running on it and perform all functions of the system application needs to support needed interfaces regarding to operating system and end device, that runs the application.

Since the system provides information transfer network, it is necessary to describe the format of information.

### 3.3. The process of launching applications

The main tasks of the application are the exchange commands and their processing. The command is transmitted using the JSON data format, which is cross-platform and allows the same identification of data regardless of the platform and programming language.

The command consists of four parts: the length of the header, header, data and an additional field. The header length field stores information about the command transmission path, the size of the

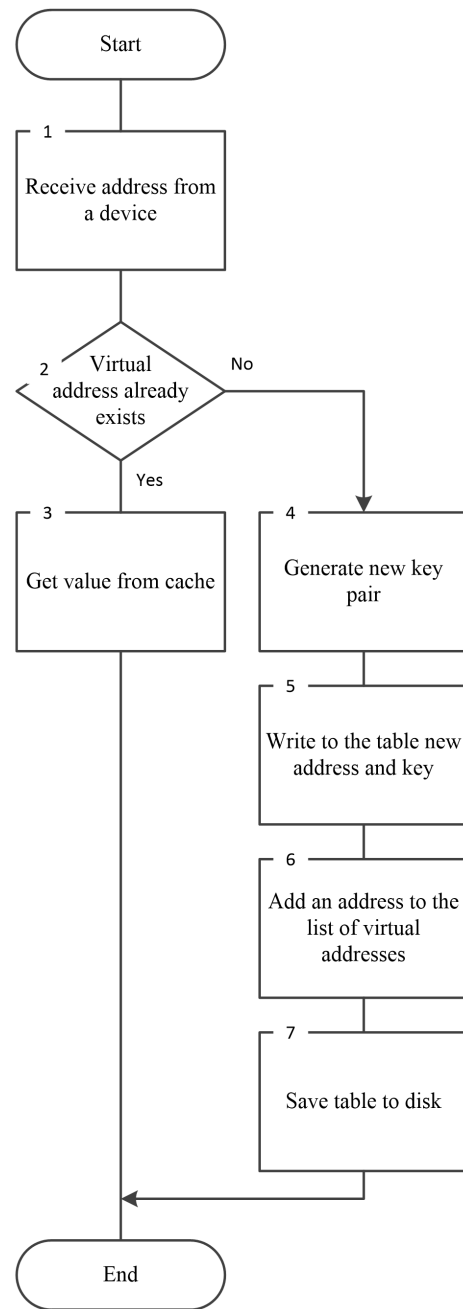
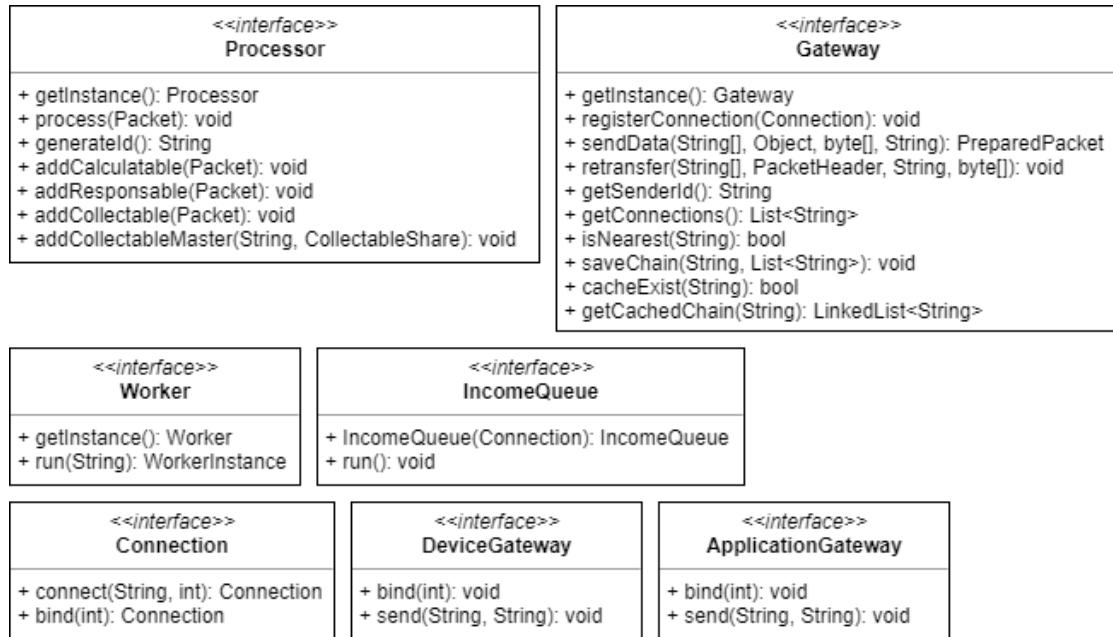


Figure 2: Process of generation key pair for devices





**Figure 3:** System class diagram

data field and the additional field in bytes, as well as information necessary for the correct processing of the request and is used for the correct delivery of the header itself. The data field stores the payload as a command if it is transmitted. Only the header and data are sent in JSON format, in plain text. The additional field stores data that cannot be expressed as text (such as binaries). Having received and correctly read all four parts of the command, the application can process it correctly.

If the command is for the current device, the text from JSON is converted to an object of the desired type. Further, depending on the object, either a specific class method is called, or the data is passed on [20, 21].

To launch additional client's applications, a special system command is used. It specifies application and status IDs, as well as initial data. Upon receiving the command, the node reads from the system the application description file, which contains the fields: tag image of the container; command to run the application; link to the first part of the application file and the hash sum of the file (optional); a list of auxiliary files and links to the initial bytes of these files and the hash sum of the file (optional). Upon receiving the file, the node checks whether the image has already been downloaded. If the image is loaded earlier – goes to the next step, otherwise - reads the image and in case of error interrupts the work and reports the error to the client. The following is the method of checking and loading the image:

```

public boolean isImageStoredLocally(String tag) {
    Process process = null;
    System.out.println("Checking tag: " + tag);
    try {
  
```

```

        process = Runtime.getRuntime().exec("docker images -q
            " + tag);
    } catch (IOException e) {
        return false;
    }
    BufferedReader bufferedReader = new BufferedReader(new
    InputStreamReader(process.getInputStream()));
    try {
        String imageId = bufferedReader.readLine();
        if (imageId == null) {
            return false;
        }
        return true;
    } catch (IOException e) {
        e.printStackTrace();
        return false;
    }
}

public boolean pullImage(String tag) {
    String command = " docker pull " + tag + " > /dev/null
    2>&1 && echo \"success\" || echo \"failed\"";
    Process process = null;
    System.out.println("Checking tag: " + tag);
    try {
        process = Runtime.getRuntime().exec("docker images -q
            " + tag);
    } catch (IOException e) {
        System.out.println("Can't execute command: " + command);
        return false;
    }
    BufferedReader bufferedReader = new BufferedReader(new
    InputStreamReader(process.getInputStream()));
    try {
        String response = bufferedReader.readLine();
        return response.equals("success");
    } catch (IOException e)
    {
        {
            e.printStackTrace();
        }
        return false;
    }
}

```

After downloading the image, the node generates a unique identifier – the startup ID. This

ID is used to interact with the application. In order to send application data, you need to know both the address of the node on which the application is running and the startup ID, because multiple copies of the application can run on different nodes for different users.

To establish a connection between an application running in isolation and an application that manages a network client, the node listens for standard output from the isolated environment. If the first line contains the required sequence, then a connection with the application is established, since it is assumed that at this time the applications are already ready to connect to each other [22, 23]. All other lines are redirected to the user who requested the execution. However, the application does not write anything to the sandbox standard input, which prevents applications from running in terminal mode. The following is a simplified source code of the method for launching an application in an isolated environment and connecting to the application that is executed:

```
if (isImageExists) {
    try {
        Process dockerProcess = docker.run(appDescriptor.container,
            appInstanceId, appDescriptor.command, this.args);
        System.out.println("==> Connecting to application");
        BufferedReader bufferedReader = new BufferedReader(new
            InputStreamReader(dockerProcess.getInputStream()));
        String line = null;
        boolean isFirstLine = true;
        while ((line = bufferedReader.readLine()) != null) {
            if (isFirstLine) {
                appSocket = new AppSocket(this);
                if (appSocket.isConnected())
                {
                    appSocket.setOnCommand(this::onCommandFromContainer);
                    appSocket.bind();
                }
                String[] args = {};
                appSocket.sendCommand("onStart", args);

                ApplicationEvent applicationEvent = new ApplicationEvent();
                applicationEvent.eventArgs = args;
                applicationEvent.eventName = "onStart";
                Gateway.getInstance()
                    .sendData(packetHeader.reverse,
                        applicationEvent, new byte[0],
                        packetHeader.targetId)
                    .send();
                isFirstLine = false;
            }
            else
```

```

        {
            response.line = line;
Gateway.getInstance()
                .sendData(packetHeader.reverse, response,
                    new byte[0], packetHeader.targetId)
                .send();
        }
    }
    System.out.println("==> App finished his work");
    docker.cleanUp(appInstanceId);
}
    catch (IOException e)
    {
        e.printStackTrace();
    }
}

```

After starting the sandbox, applications have access to almost all the capabilities of the end device on which they run. For example, an application has access to multithreading, RAM and memory, video memory and video adapter, to the network itself and devices connected to it. However, at the same time, the application does not have access to the rest of the operating system, i.e. it cannot receive anything outside its area. It can see the total amount of used memory, but it cannot see what exactly the memory is occupied. In this case, it is impossible to access the user's files.

It is also impossible to get information on connections outside the isolated environment, but it is possible within the isolated part. In fact, the application works in a virtual area, but without a virtualization layer and therefore does not have performance limitations in comparison with conventional virtual machines, but, if necessary, can use any capabilities of the operating system. In addition, it is possible to install third-party software, which is subsequently removed along with the removal of the sandbox and all data that is generated by the application and which is not stored in distributed memory or another area.

### 3.4. Applying and testing the system

Research was carried out on a Mac mini. ESP8266 and ESP32 boards were used as devices, which were connected to a computer using a WiFi network. For testing, a program developed that calculates the factorial of large numbers (figure 4).

The developed system can be applied wherever there may be any calculations. Using this system, low-power devices can request heavy computations, which significantly expands the list of tasks that these devices can solve.

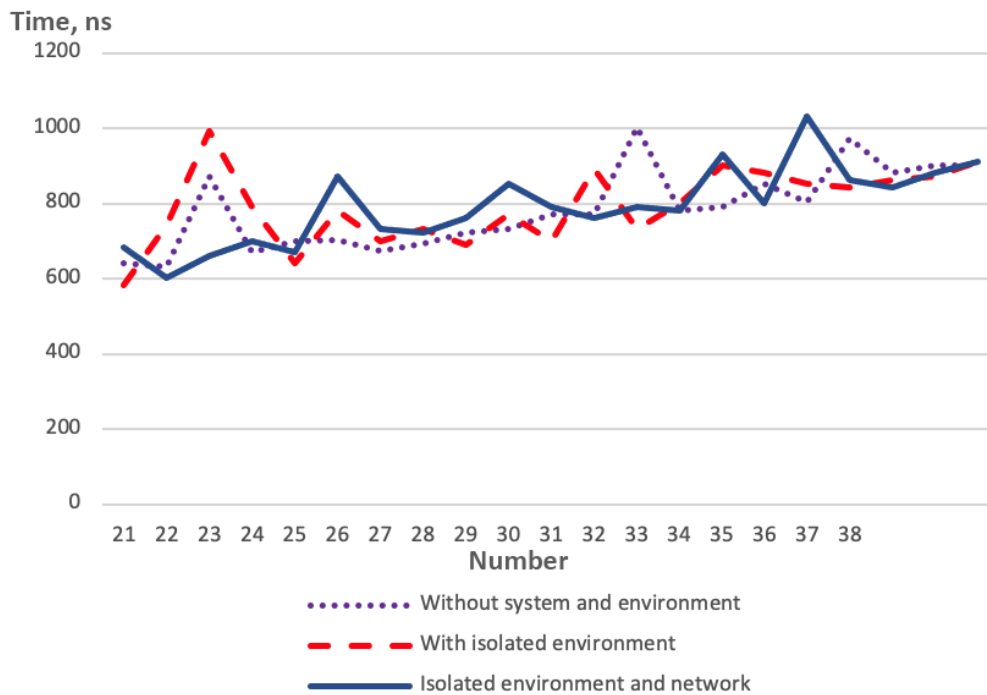


Figure 4: Comparing calculation speed

## 4. Conclusions

In this article improved the devices and applications interaction process in distributed systems, a model of a distributed system is created, and implemented a software protocol to support the interaction of network devices and applications. Methods of interaction of system parts and application components are developed.

The security of data transmission and their processing are one of the key in modern systems and data transmission networks, therefore, the greatest attention is paid to solving this problem in the process of implementing the system and algorithms of its operation. Application of the developed encryption and data transmission methods are increasing security of data transmission due to multi-layer encryption, and the use of an isolated environment makes the process of executing client applications independent of the host operating system and other applications that can be launched and can interfere with or intercept data. In the process of further research, it is planned to expand the capabilities of the system, namely to add support for graphical applications, parallelize processes on network devices and implement more functionality for interacting with the system and IoT devices.

## References

- [1] A. Amer, N. Abu, Comparison study between ipv4 & ipv6, *International Journal of Computer Science Issues (IJCSI)* 9 (2012) 314–317.
- [2] G. Kirichek, V. Tymoshenko, O. Rudkovskiy, S. Hrushko, Decentralized system for run services, in: *Proceedings of the Second International Workshop on Computer Modeling and Intelligent Systems, CMIS '2019, CEUR Workshop Proceedings 2353, Zaporizhzhia, Ukraine, 2019*, pp. 860–872.
- [3] G. Kirichek, S. Skrupsky, M. Tiahunova, A. Timenko, Implementation of web system optimization method, in: *Proceedings of the Third International Workshop on Computer Modeling and Intelligent Systems, CMIS '2020, CEUR Workshop Proceedings 2608, Zaporizhzhia, Ukraine, 2020*, pp. 199–210.
- [4] W. Felter, A. Ferreira, R. Rajamony, J. Rubio, An updated performance comparison of virtual machines and linux containers, in: *2015 IEEE international symposium on performance analysis of systems and software, ISPASS '2015, IEEE, Philadelphia, PA, USA, 2015*, pp. 171–172. doi:10.1109/ISPASS.2015.7095802.
- [5] T. Mauro, Adopting Microservices at Netflix: Lessons for Architectural Design, 2015. [https://www.https://www.nginx.com/blog/microservices-at-netflix-architectural-best-practices/](https://www.nginx.com/blog/microservices-at-netflix-architectural-best-practices/).
- [6] A. Gencer, S. Basu, I. Eyal, R. V. Renesse, E. Sirer, Decentralization in bitcoin and ethereum networks, in: *International Conference on Financial Cryptography and Data Security, FC '2018, Springer, Berlin, Heidelberg, 2018*, pp. 439–457. doi:10.1007/978-3-662-58387-6\_24.
- [7] A. Antonopoulos, G. Wood, *Mastering Ethereum: Building Smart Contracts and DApps*, O'Reilly Media, Inc, 2018.
- [8] G. Kirichek, V. Kurai, Implementation quadtree method for comparison of images, in: *14th International Conference on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering, TCSET '2018, IEEE, Slavske, Ukraine, 2018*, pp. 129–132. doi:10.1109/TCSET.2018.8336171.
- [9] M. Kumar, *Energy Efficiency of Java Programming Language*, Wayne State University, 2018.
- [10] D. Abts, *Masterkurs Client/Server-Programmierung mit Java: Anwendungen entwickeln mit Standard-Technologien*, Springer-Verlag, 2019.
- [11] C. P. Mayer, Security and privacy challenges in the internet of things, *Electronic Communications of the EASST* 17 (2009) 1–12. doi:10.14279/tuj.eceasst.17.208.
- [12] W. Jansen, T. Grance, Guidelines on security and privacy in public cloud computing, in: *NIST Special Publication 800-144, U.S. Department of Commerce, NIST Pubs, 2011*.
- [13] N. Pathania, Setting up jenkins on docker and cloud, in: *Pro Continuous Delivery*, Apress, Berkeley, CA, 2017, pp. 115–143.
- [14] M. Miraz, M. Ali, P. Excell, R. Picking, Internet of nano-things, things and everything: future growth trends, in: *Future Internet*, volume 10, 2018.
- [15] G. Kirichek, D. Kyrychek, S. Hrushko, A. Timenko, Implementation the protection method of data transmission in network, in: *IEEE International Conference on Advanced Trends in Information Theory, ATIT '2019, IEEE, Kyiv, Ukraine, 2019*, pp. 129–132. doi:10.1109/

ATIT49449.2019.9030482.

- [16] A. Corbellini, Elliptic curve cryptography: a gentle introduction, 2015.
- [17] B. Gulmezoglu, M. Inci, G. Irazoqui, T. Eisenbarth, B. Sunar, Cross-vm cache attacks on aes, in: IEEE Transactions on Multi-Scale Computing Systems, volume 10, IEEE, 2016, pp. 211–222. doi:10.1109/TMSCS.2016.2550438.
- [18] S. Kumark, M. Girimondo, A. Weimerskirch, C. Paar, A. Patel, A. Wander, Embedded end-to-end wireless security with ecdh key exchange, in: IEEE 2003 46th Midwest Symposium on Circuits and Systems, volume 2, IEEE, Cairo, Egypt, 2003, pp. 786–789. doi:10.1109/MWSCAS.2003.1562404.
- [19] G. Kirichek, V. Harkusha, A. Timenko, N. Kulykovska, System for detecting network anomalies using a hybrid of an uncontrolled and controlled neural network, in: Proceedings of the Computer Science & Software Engineering: Proceedings of the 2nd Student Workshop, CS&SE@SW '2019, CEUR Workshop Proceedings 2546, Kryvyi Rih, Ukraine, 2019, pp. 138–148.
- [20] S. Sriparasa, JavaScript and JSON essentials, Packt Publishing, 2013.
- [21] F. Pezoa, J. Reutter, F. Suarez, M. Ugarte, D. Vrgoč, Foundations of json schema, in: Proceedings of the 25th International Conference on World Wide Web, WWW '16, 2016, pp. 263–273. doi:10.1145/2872427.2883029.
- [22] D. Merkel, Docker: lightweight linux containers for consistent development and deploymentno-things, things and everything: future growth trends, in: Linux journal, volume 2014, 2014, pp. 1–5.
- [23] T. Bui, Analysis of docker security, in: Aalto University T-110.5291 Seminar on Network Security, arXiv preprint arXiv: 1501.02967v1, 2015.

# Coverage of the coronavirus pandemic through entropy measures

Vladimir N. Soloviev<sup>a,b</sup>, Andrii O. Bielinskyi<sup>a</sup> and Natalia A. Kharadzjan<sup>a</sup>

<sup>a</sup>Kryvyi Rih State Pedagogical University, 54 Gagarin Ave., Kryvyi Rih, 50086, Ukraine

<sup>b</sup>Bohdan Khmelnytsky National University of Cherkasy, 81 Shevchenko Blvd., Cherkasy, 18031, Ukraine

## Abstract

The rapidly evolving coronavirus pandemic brings a devastating effect on the entire world and its economy as a whole. Further instability related to COVID-19 will negatively affect not only on companies and financial markets, but also on traders and investors that have been interested in saving their investment, minimizing risks, and making decisions such as how to manage their resources, how much to consume and save, when to buy or sell stocks, etc., and these decisions depend on the expectation of when to expect next critical change. Trying to help people in their subsequent decisions, we demonstrate the possibility of constructing indicators of critical and crash phenomena on the example of Bitcoin market crashes for further demonstration of their efficiency on the crash that is related to the coronavirus pandemic. For this purpose, the methods of the theory of complex systems have been used. Since the theory of complex systems has quite an extensive toolkit for exploring the nonlinear complex system, we take a look at the application of the concept of entropy in finance and use this concept to construct 6 effective entropy measures: Shannon entropy, Approximate entropy, Permutation entropy, and 3 Recurrence based entropies. We provide computational results that prove that these indicators could have been used to identify the beginning of the crash and predict the future course of events associated with the current pandemic.

## Keywords

coronavirus, Bitcoin, cryptocurrency, crash, critical event, measures of complexity, entropy, indicator-precursor

## 1. Introduction

The novel coronavirus outbreak (COVID-19) quickly became a catastrophic challenge for the whole world, and it would be even more of a “black swan” than the global financial crisis and Great recession of 2008-2009 [1]. While the governments are focused on saving lives and preventing the virus from spreading further, at the moment, it continuously led to a substantial disruption of the world financial system. Furthermore, economic worries are also increasing dramatically. At first, people have worried about potentially become one of the infected. Then,

---

*CS&SE@SW 2020: 3rd Workshop for Young Scientists in Computer Science & Software Engineering, November 27, 2020, Kryvyi Rih, Ukraine*

✉ vnsoloviev2016@gmail.com (V.N. Soloviev); krivogame@gmail.com (A.O. Bielinskyi);

n.a.kharadzjan@gmail.com (N.A. Kharadzjan)

🌐 <https://kdpu.edu.ua/personal/vmsoloviov.html> (V.N. Soloviev);

<https://kdpu.edu.ua/personal/nakharadzhian.html> (N.A. Kharadzjan)

🆔 0000-0002-4945-202X (V.N. Soloviev); 0000-0002-2821-2895 (A.O. Bielinskyi); 0000-0001-9193-755X (N.A.

Kharadzjan)



© 2020 Copyright for this paper by its authors.  
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)



when the situation escalated, people started to worry about the impact on their economic situation [2, 3, 4]. To prevent exponential growth of the disease incidence bans on traveling the world, visiting public places, and businesses were established that had particularly a destructive impact on financial markets, and the cryptocurrency market, showing that for most systems it was unexpected. Subsequent days on the market promise to be more volatile and riskier than before. And it seems that Bitcoin is going to fall into a new Great recession along with the most influential stock indexes. As it was noted by Pier Francesco Procacci, Carolyn E. Phelan and Tomaso Aste [5], the crisis-state representative and eventually becomes extremely dominant in March, but a lot of factors have to be included to make an adequate analysis of this problem.

The doctrine of the unity of the scientific method states that for the study of events in socio-economic systems, the same methods and criteria as those used in the study of natural phenomena are applicable. The increasing mathematical knowledge of complex structure provides us with such methods and tools that are interconnected and can be used to study so relevant and “almost unpredictable” situations as the coronavirus pandemic. By almost unpredictable, we mean that although past experience and data are not so relevant in this case, we can use some mathematical methods (models) that are useful in current situations, including current statistics, relevant parameters, and even the inner complexity of the system. As an example, Alexis Akira Toda [6] estimated the Susceptible-Infected- Recovered (SIR) epidemic model for COVID-19 because of its desire to help individuals in managing their investments. Governments also need help in making an informed decision on imposing travel restrictions, social distancing, closure of schools and businesses, etc., and, mainly, for how long [7]. In the paper of Bohdan M. Pavlyshenko [8] was studied different regression approaches for modeling COVID-19 spread and its impact on the stock market. The logistic curve model was used under Bayesian regression for predictive analytics of the coronavirus spread. The obtained results showed that different crises with different reasons have a different impact on the same stocks. Bayesian inference makes it possible to analyze the uncertainty of crisis. Michele Costola, Matteo Iacopini and Carlo R. M. A. Santagiustina [9] regarding the public concern of 6 countries during the outbreak of COVID-19 find that the Italian index in the current situation is relevant in explaining index returns for other studied countries. Assuming that COVID-19 has a deterministic, exogenous impact on the market, Karina Arias-Calluari, Fernando Alonso-Marroquin, Morteza Nattagh-Najafi and Michael Harré [10] forecast it with a model of the stochastic and systematic risk. Here, such a model is assumed to be  $q$ -Gaussian diffusion process which is accompanied by three spatio-temporal regimes. In this case, the results were achieved with 85% accuracy. Presented results are promising not only for markets but also for risk control in other areas such as seismology, communication networks, etc. In the paper [11] the bandwidth and the discount factor are proposed to minimize a criterion consistent with the traditional requirements of the validation of a probability density forecast. They use Kolmogorov-Smirnov statistic and a discrepancy statistic to build a quantitative criterion of the accuracy of pdf which they try to maximize when selecting the bandwidth and the discount factor of their time-varying pdf. Such an approach allows exposing an accurate chronology of the current pandemic. Ayoub Ammy-Driss and Matthieu Garcin [12] explore novel pandemic using two efficiency indicators: the Hurst exponent and the memory parameter of a fractional Lévy-stable motion. Presented results highlight the occurrence of inefficiency at the almost beginning of the crisis for US indices. Asian and Australian indices are seemed to be less affected

during this period, i.e., their inefficiency is even questionable. A. Fronzetti Colladon, S. Grassi, F. Ravazzolo and F. Violante [13] present a new textual data index, which assesses the related most usable keywords and semantic network position, for predicting stock market data. They apply it to the Italian press and use it for predicting recent periods of Italian stock, including the COVID-19 crisis. According to the results, it can be observed that the index is characterized by strong predictability. Unfortunately, by the luck of observation and potentially diseased people that are surrounding us at the moment, it is still difficult to predict further dynamic during the pandemic.

As it can be observed, markets have seen significant numbers of investors selling off and rebalancing their portfolios with less risky assets. That has been leading to large losses and high volatilities, typical of crisis periods. The economy key for preventing such activity may lie in cryptocurrency and constructing effective indicators of possible critical states that will help investors and traders fill in safety. Bitcoin, which is associated with the whole crypto market, has such properties as detachment and independence from the standard financial market and the proclaimed properties that should make it serve as the 'digital gold' [14]. As was shown by Ladislav Kristoufek [15], Bitcoin promises to be a safe-haven asset with its low correlation with gold, S&P 500, Dow Jones Industrial Average, and other authoritative stock indices even in the extreme events. But authors please not overestimate the cryptocurrency since according to their calculations and, obviously, the current structure of the system, gold remains more significant. But for ten years, this token has been discussed by many people, it has experienced a lot in such a short period, many people believe in it, and it has managed to form a fairly complex and self-organized system. The integrated actions from real-world merge in such dynamics and relevant information that is encoded in Bitcoin's time series can be extracted [16, 17, 18]. In the context of volatile financial markets, it is important to select such measures of complexity that will be able to notify us of upcoming abnormal events in the form of crises at an early stage.

In this article, we:

- present such measures;
- study critical and crash phenomena that have taken place in the cryptocurrency market;
- try to understand whether a crash caused by the coronavirus pandemic could have been identified and predicted by such informative indicators or not.

According to our goals and actions, the paper is structured as follows. In Section 2, we presented a brief overview of the studies in this field of science. In Section 3, relying on these researches and the experience of other scientists, we present our classification of Bitcoin's crises for the period from 1 January 2013 to 9 April 2020. In Section 4, we describe the applied methods and present some empirical results with their subsequent description. Section 5 concludes.

## 2. Review of the previous studies

For its short history of existence, bitcoin has experienced many events, periodic rises and sudden dips in specific periods, regulatory actions, and discussions about whether it can become a universally mature commodity around the world or not, and therefore its largely unexplored dynamics is still presenting new opportunities and challenges for traders, economists, and researchers from different fields of science to highlight chaos that is hidden in it. Although officially bitcoin is considered to be a commodity rather than a currency, the comprehensive analysis of the applicability of the instruments which have been used for mature financial markets for a long time can be made for cryptocurrencies and, namely, for the Bitcoin market.

A vast amount of different methods, as an example, from the theory of complexity, the purpose of which is to quantify the degree of complexity of systems obtained from various sources of nature, can be applied in our study. Such applications have been studied intensively for an economic behavior system. As an example, Miguel Henry and George Judge [19] used an information theoretic-symbolic logic approach which is based on Shannon's information entropy and called *Permutation entropy* (PE<sub>n</sub>). The entropy is applied to the Dow Jones Industrial Average to extract information from this complex economic system. The result demonstrates the ability of the PE<sub>n</sub> method to detect the degree of disorder and uncertainty for the specific time that is explored. In such paper, [20] presented by Higor Sigaki, Matjaž Perc, and Haroldo Valentin Ribeiro, the PE<sub>n</sub> and statistical complexity over sliding time-window of daily closing price log-returns are used to quantify the dynamic efficiency of more than four hundred cryptocurrencies. Authors address to the efficient market hypothesis when the values of two statistical measures within a time-window cannot be distinguished from those obtained by chance. They find that 37% of the cryptocurrencies in their study stay efficient over 80% of the time, whereas 20% are informationally inefficient in less than 20% of the time. Moreover, the market capitalization is not correlated with their efficiency. Performed analysis of information efficiency over time reveals that different currencies with similar temporal patterns form four clusters, and it is seen that more young currencies tend to follow the trend of the most leading currencies.

Steve M. Pincus and Rudolf E. Kalman [21], considering both empirical data and models, including composite indices, individual stock prices, the random-walk hypothesis, Black-Scholes, and fractional Brownian motion models to demonstrate the benefits of *Approximate entropy* (ApEn), a quantitative measure of sequential irregularity. On the example of the applied models and empirical data, the authors presented that ApEn can be readily applied to the classical econometric modeling apparatus. This research the usefulness of ApEn on the example of three major events of the stock market crash in the US, Japan, and India. During the major crashes, there is significant evidence of a decline of ApEn during and pre-crash periods. Based on the presented results, their research concludes that ApEn can serve as a base for a good trading system. This article [22] gives evidence of the usefulness of Approximate Entropy. The researchers quantify the existence of patterns in evolving data series. In general, scientists observe that the degree of predictability increases in times of crisis. However, the presented results do not demonstrate that the regularity techniques studied in this paper can serve to predict an imminent crash.

Also, there is a paper [23] that is dedicated to Ethereum. Here, the concept of entropy is

applied for characterizing the nonlinear properties of the cryptocurrencies. For their goal, Shannon, Tsallis, Rényi, and Approximate entropies are estimated. From their empirical results, it is obtained that all entropies are positive. Of great interest is the results of ApEn which demonstrates larger value for Ethereum than for Bitcoin. In this case, it concludes that Ethereum has higher volatility. The same result for other measures. Daniel Traian Pele and Miruna Mazurencu [24] investigate the ability of several econometrical models to forecast value at risk for a sample of daily time series of cryptocurrency returns. Using high-frequency data for Bitcoin, they estimate the entropy of the intraday distribution of log-returns through the symbolic time series analysis (STSA), producing low-resolution data from high-resolution data. Their results show that entropy has strong explanatory power for the quantiles of the distribution of the daily returns. They confirm the hypothesis that there is a strong correlation between the daily logarithmic price of Bitcoin and the entropy of intraday returns. Based on Christoffersen's tests for Value at Risk (VaR) backtesting, they conclude that the VaR forecast built upon the entropy of intraday returns is the best, compared to the forecasts provided by the classical GARCH models.

During our investigation, we found that many theses and papers that have been studied the dynamics of financial markets rather than cryptocurrencies using the measures from the theory of complexity. Thus, we are interested in contributing to the study of the dynamics of a potentially strong currency that affects the world economy and faces the same problems as even the most quoted financial markets. Thus, the construction of predictive models and measures of unexpected and critical events in the cryptocurrency market remains relevant.

### **3. Data preparation and classification**

The ecosystem of cryptocurrencies has been growing at an increasing pace. More and more of them become tradable and begin to inspire confidence for many people. Such events as the coronavirus threat that are presented to be unpredictable, break this confidence and the subsequent anxiety of traders shape the extent of economic worries and collapse.

Previously, we conducted research on the dynamics of both stock and cryptocurrency markets where we provided results confirming the effectiveness of using quantitative methods of the theory of complexity which can serve as a base for estimation of indicators-precursors of the critical states. This paper presents a comparative analysis of the measures based on the concept of entropy: Shannon, Approximate, Sample, Permutation, and Recurrence based entropies. Here, each entropy is applied to the entire market and separate bubble, crashes, and critical events that have taken place in this market. Thus, we advanced into action and set the tasks:

- Classification of such bubbles, critical events and crashes.
- Construction of such indicators that will predict crashes, critical events in order to give investors and ordinary users the opportunity to trade in this market.

At the moment, there are various research papers on what crises and crashes are and how to classify such interruptions in the market of cryptocurrencies. Taking into account the experience of previous researchers and our own [25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36],

we present our classification of such leaps and falls from the previous articles [28, 29, 30, 33] on the cryptocurrency topic, relying on Bitcoin time series during the period (01.01.2013 – 09.04.2020) of verifiable fixed daily values of the Bitcoin price (BTC) (<https://finance.yahoo.com/cryptocurrencies>).

For our classification, crashes are short, time-localized drops, with the strong losing of price per each day, which are formed as a result of the bubble. Critical events are those falls that could go on for a long period, and at the same time, they were not caused by a bubble. The bubble is increasing in the price of the cryptocurrency that could be caused by certain speculative moments. Therefore, according to our classification of the event with the number (1-3, 6-8, 12-15, 17) are the crashes, all the rest - critical events. More detailed information about crises, crashes, and their classification following these definitions are given in table 1.

**Table 1**  
BTC HISTORICAL CORRECTIONS. LIST OF BITCOIN MAJOR CORRECTIONS  $\geq 20\%$  SINCE APRIL 2013

№	Name	Days in correction	Bitcoin High Price, \$	Bitcoin Low Price, \$	Decline, %	Decline, \$
1	08.04.2013-15.04.2013	8	230.00	68.36	70	161.64
2	04.12.2013-18.12.2013	15	1237.66	540.97	56	696.69
3	05.02.2014-25.02.2014	21	904.52	135.77	85	768.75
4	12.11.2014-14.01.2015	64	432.02	164.91	62	267.11
5	26.01.2015-31.01.2015	5	269.18	218.51	20	50.67
6	09.11.2015-11.11.2015	3	380.22	304.70	20	75.52
7	18.06.2016-21.06.2016	4	761.03	590.55	22	170.48
8	04.01.2017-11.01.2017	8	1135.41	785.42	30	349.99
9	03.03.2017-24.03.2017	22	1283.30	939.70	27	343.60
10	10.06.2017-15.07.2017	36	2973.44	1914.08	36	1059.36
11	31.08.2017-13.09.2017	13	4921.85	3243.08	34	1678.77
12	16.12.2017-22.12.2017	7	19345.49	13664.9	29	5680.53
13	13.11.2018-26.11.2018	13	6339.17	3784.59	40	2554.58
14	09.07.2019-16.07.2019	7	12567.02	9423.44	25	3143.58
15	22.09.2019-29.09.2019	7	10036.98	8065.26	20	1971.7
16	27.10.2019-24.11.2019	28	9551.71	7047.92	26	2503.79
17	06.03.2020-16.03.2020	11	9122.55	5014.48	45	4108.07

Accordingly, during this period in the Bitcoin market, many crashes and critical events shook it. Thus, considering them, we emphasize 17 periods on Bitcoin time series, whose falling we predict by our indicators, relying on normalized returns, where usual returns are calculated as:

$$G(t) = \ln x(t + \delta t) - \ln x(t) \cong [x(t + \delta t) - x(t)] / x(t)$$

and normalized returns as:

$$g(t) \cong [G(t) - \langle G \rangle] / \sigma,$$

where  $\sigma$  is a standard deviation of  $G$ ,  $\delta t$  is a time lag (here  $\delta t = 1$ ), and  $\langle \dots \rangle$  denotes the average over the period under study.

Further calculations were carried out within the framework of the algorithm of a moving window. For this purpose, the part of the time series (window), for which there were calculated measures of complexity, was selected, then the window of a length 100 was displaced along with the time series in a one-day increment and the procedure repeated until all the studied series had exhausted. Further, comparing the dynamics of the actual time series and the corresponding measures of complexity, we can judge the characteristic changes in the dynamics of the behavior of complexity with changes in the cryptocurrency. The key idea here is the hypothesis that the complexity of the system before the crashes and the actual periods of crashes must change. This should signal the corresponding degree of complexity if they are able to quantify certain patterns of a complex system. A significant advantage of the introduced measures is their dynamism, that is, the ability to monitor the change in time of the chosen measure and compare it with the corresponding dynamics of the output time series. This allowed us to compare the critical changes in the dynamics of the system, which is described by the time series, with the characteristic changes of concrete measures of complexity. It turned out that quantitative measures of complexity respond to critical changes in the dynamics of a complex system, which allows them to be used in the diagnostic process and prediction of future changes.

Moreover, using the measures already mentioned, we are trying to understand whether certain patterns or information were hidden in the market that could have enabled us to predict the current crisis that was caused by the coronavirus pandemic, or not. Therefore, in the next section, we give a brief description of the applied methods and present the empirical results.

For further analysis, the computations in a rolling window algorithm with a length of 100 and a step of 1 for the entire time series and local crashes we made. From the classification table, we selected 4 crashes with numbers 1, 3, and 7. This choice was due to the fact that the current crisis is essentially a crash according to our classification, and therefore for convenient presentation of complexity measures in further, we should select a few of any crashes, the nature of which in some sense would be similar to the main crash of the presented article. For the entire Bitcoin time series, each crash and the critical event will be marked by the arrow in accordance with the table. For local crashes, the beginning of each crash has to be indicated by our measures in the time of 100.

## 4. Related methods

Nowadays, the most important quantity that allows us to parameterize complexity in deterministic or random processes is entropy. Originally, it was introduced by Rudolf Clausius [37], in the context of classical thermodynamics, where according to his definition, entropy tends to increase within an isolated system, forming the generalized second law of thermodynamics. Then, the definition of entropy was extended by Boltzmann and Gibbs (BG) [38, 39], linking it to molecular disorder and chaos, to make it suitable for statistical mechanics, where they combined the notion of entropy and probability [40].

After the fundamental paper of Claude Elwood Shannon [41] in the context of information theory, its notion was significantly redefined. After this, it has been evolved along with different ways and successful enough used for the research of economic systems [42, 43, 44]. In

what follows, the theoretical background and empirical results for the most popular entropy measures are presented.

#### 4.1. Shannon entropy

*Shannon entropy* (ShEn) was proposed as a measure of uncertainty by its author – Claude Elwood Shannon in his famous paper “A Mathematical Theory of Communication” [41], where, at first, its purpose was to quantify the degree of ‘lost information’ in phone-line signals. His contribution has proved that this approach can be generalized for any series where probabilities exist. Comparatively to the entropy of Clausius and Boltzmann that were valid only for thermodynamic systems, it was significant progress. Formally, his approach can be briefly defined as the average amount of ‘information’ and ‘uncertainty’ encoded in patterns recorded from a signal, or message. Other interpretations refer to entropy as a measure of ‘chaos’ or disorder in a system. During decades the generalization of Shannon’s entropy has been applied to various domains, particularly to the financial sector.

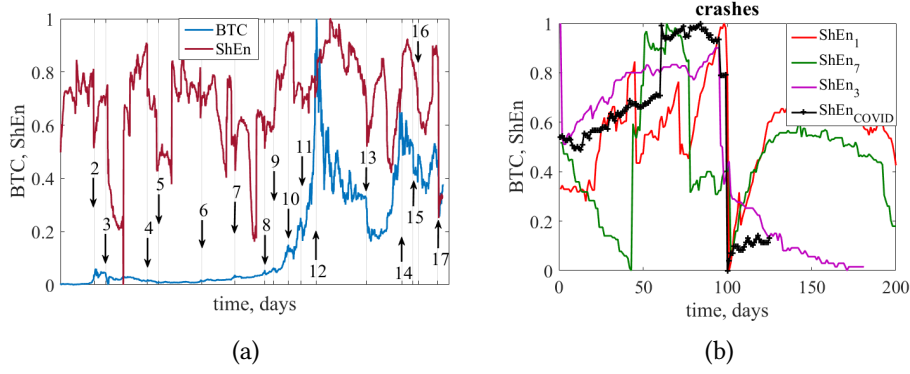
The general approach can be described as follows. Formally, we represent the underlying dynamic state of the system in probability distribution form  $P$  and then ShEn  $S$  with an arbitrary base (i.e. 2,  $e$ , 10) is defined as:

$$S [P] = - \sum_{i=1}^N p_i \log p_i, \quad (1)$$

where  $p_i$  represents the probability that price  $i$  occurs in the sample’s distribution of the Bitcoin time series, and  $N$  is the total amount of data in our system. When dealing with continuous probability distributions with a density function  $f(x)$ , we can define the entropy as:

$$H (f) = - \int_{-\infty}^{+\infty} f(x) \log f(x). \quad (2)$$

According to the approach, the negative log increases with rarer events due to the information that is encoded in them (i.e. they surprise when they occur). Thus, when all  $p_i$ ’s have the same value, i.e. where all values are equally probable, and  $S [P]$  reaches its minimum for more structured time series (events that are more certain). Equation 2 is obeyed to the same rules as discrete version of this method. In the figure 1 are the empirical results for ShEn, for the entire time series (1a), and local crashes (1b).



**Figure 1:** ShEn dynamics along with the entire time series of Bitcoin (a). The dynamics of ShEn for the local crashes (b) in accordance with the table 1

## 4.2. Approximate entropy

To gain more detail analysis of the complex financial systems, it is known other entropy methods have become known, particularly, ApEn developed by Steve M. Pincus [45] for measuring regularity in a time series.

When calculating it, given  $N$  data points  $\{x(i) \mid i = 1, \dots, N\}$  are transformed into subvectors  $\vec{X}(i) \in \mathcal{R}^{d_E}$ , where each of those subvectors has  $[x(i), x(i+1), \dots, x(i+d_E-1)]$  for each  $i, 1 \leq i \leq N-d_E+1$ . Correspondingly, for further estimations, we would like to calculate a probability of finding such patterns whose Chebyshev distance  $d[\vec{X}(i), \vec{X}(j)]$  does not exceed a positive real number  $r$ :

$$C_i^{d_E}(r) = (N - d_E + 1)^{-1} \sum_{j=1}^{N-d_E+1} \mathcal{H}(r - d[\vec{X}(i), \vec{X}(j)])$$

where  $\mathcal{H}(\cdot)$  is the Heviside function which count the number of instances  $d[\vec{X}(i), \vec{X}(j)] \leq r$ .

Next, we estimate

$$F^{d_E}(r) = (N - d_E + 1)^{-1} \sum_{i=1}^{N-d_E+1} \ln(C_i^{d_E}(r)),$$

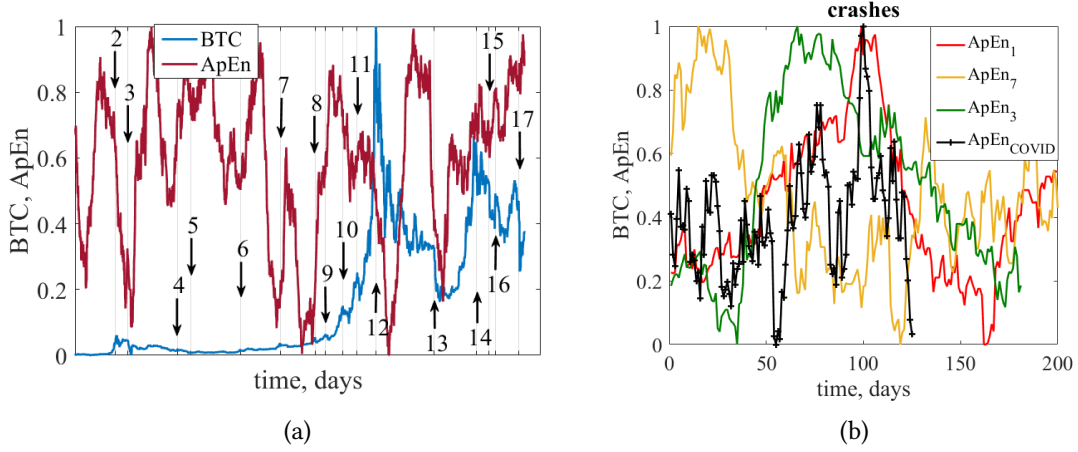
and ApEn of a corresponding time series (for fixed  $d_E$  and  $r$ ) measures the logarithmic likelihood that patterns that are close for  $d_E$  adjacent observations remain close on the next comparison:

$$ApEn(d_E, r, N) = F^{d_E}(r) - F^{d_E+1}(r). \quad (3)$$

If equation (3) is small, then we should expect high regularity in our data (sequences remain close to each other), and extreme value of ApEn indicates independent sequential processes.

The calculation results for the full time series (2a) and the local ApEn values (2b) are presented in figure 2.





**Figure 2:** ApEn dynamics along with the entire time series of Bitcoin (a). The dynamics of ApEn for the local crashes (b) in accordance with the table 1

### 4.3. Permutation entropy

PEn, according to the previous approach, is a complexity measure that is related to the fundamental Information theory and entropy proposed by Shannon. Such a tool was proposed by C. Bandt and B. Pompe [46], which is characterized by its simplicity, computational speed that does not require some prior knowledge about the system, strongly describes nonlinear chaotic regimes. Also, it is characterized by its robustness to noise [47, 48] and invariance to nonlinear monotonous transformations [49]. The combination of entropy and symbolic dynamics turned out to be fruitful for analyzing the disorder for the time series of any nature without losing their temporal information. According to this method, we need to consider “ordinal patterns” that consider the order among time series and relative amplitude of values instead of individual values. For evaluating PEn, at first, we need to consider a time series  $\{x(i) \mid i = 1, \dots, N\}$  which relevant details can be “revealed” in  $d_E$ -dimensional vector

$$\vec{X}(i) = [x(i), x(i + \tau), \dots, x(i + (d_E - 1)\tau)],$$

where  $i = 1, 2, \dots, N - (d_E - 1)\tau$ , and  $\tau$  is an embedding delay of our time delayed vector. After it, we map by the ordinal pattern, related to time  $t$ , we consider the permutation pattern  $\pi_l(t) = (k_0, k_1, \dots, k_{d_E-1})$  where  $1 \leq l \leq m!$  if the following condition is satisfied:

$$x(j + k_0\tau) \leq x(j + k_1\tau) \leq \dots \leq x(j + k_{d_E-1}\tau). \quad (4)$$

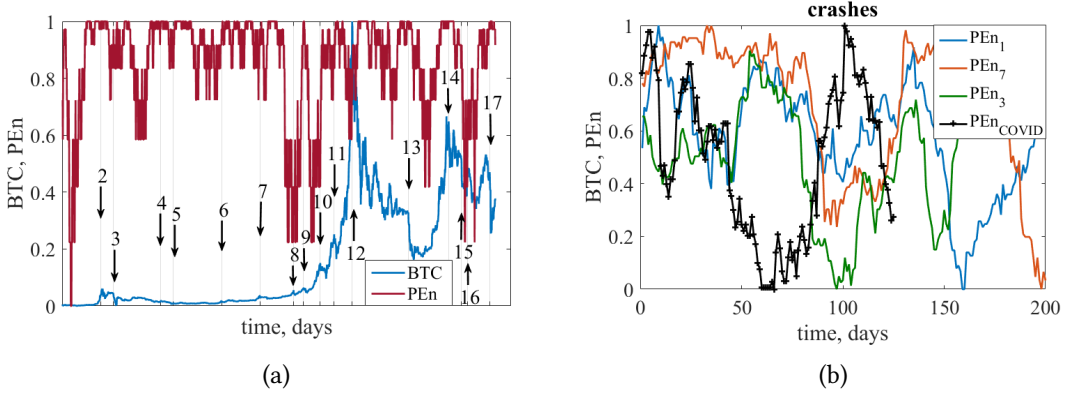
Then, regarding the probability distribution  $P$  of each ordinal pattern, we finally define the *normalized permutation entropy* as:

$$E_s[P] = \frac{-\sum_{l=1}^{d_E!} p_l \ln p_l}{\ln d_E!}, \quad (5)$$

where  $p_l$  is the relative frequency of each ordinal pattern.

The same idea has the permutation entropy. With the much lower entropy value, we get a more predictable and regular sequence of the data. Therefore, PEn gives a measure of the departure of the time series from a complete noise and stochastic time series.

There must be predefined appropriate parameters on which PEn relying, namely, the embedding dimension  $d_E$  is a paramount of importance because it determines  $d_E$  possible states for the appropriate probability distribution. With small values such as 1 or 2, parameter  $d_E$  will not work because there are only few distinct states. Furthermore, for obtaining reliable statistics and better detecting the dynamic structure of data,  $d_E$  should be relevant to the length of the time series or less [50]. For our experiments,  $d_E \in \{3, 4\}$  and  $\tau \in \{2, 3\}$  indicate the best results. Hence, in figure 3 we can observe the empirical results for PEn where it serves as indicator-precursor of the possible unusual states.



**Figure 3:** PEn dynamics along with the entire time series of Bitcoin (a). The dynamics of PEn along with the local crashes (b) in accordance with the table 1

#### 4.4. Recurrence based entropies

The corresponding measure of entropy is related to the recurrence properties that may be peculiar for the nonlinear complex system. A method that allows us to visualize and understand the recurrence behavior of the system is the *recurrence plot* (RP). A recurrence plot reflects the binary similarities of pairs of vectors in phase space. In fact, we visually represent the reconstructed  $d_E$ -dimensional phase space of the system according to *Takens' theorem* [51] by a square similarity matrix that can be defined as:

$$R_{i,j} = \mathcal{H}(\epsilon - \|\vec{X}(i) - \vec{X}(j)\|),$$

where  $i, j = 1, \dots, N$ ;  $\epsilon$  is a threshold that determines the similarity of two probably neighborhood trajectories, and previously mentioned  $\mathcal{H}$  is a Heaviside function that represents an answer in binary form;  $\|\cdot\|$  is the Chebyshev distance, and  $N$  is the size of the analyzed data. If we keep the fixed radius condition and use  $L_\infty$ -norm, as a result, the binary matrix captures a total  $N^2$  similarity values. Further recurrence plot is the representation of similar vector pairs that are represented by black dots, whereas cells referring to dissimilar pairs of vectors are presented by white dots.

Such representation of the systems can be, certainly, useful. But it has a limitation. As an example, with the increasing size of a plot, it is hard to display it graphically as a whole. It has to be resized, which in turn will lead to new artifacts and distortion of the patterns. These types of plots may cause incorrect interpretation.

However, to enable an objective assessment, the graphical representation of RP allows us to derive qualitative characterizations of the dynamical systems within a recurrence plot. For the quantitative description of the dynamics, the small-scale patterns in the RP can be used, such as diagonal and vertical lines. The histograms of the lengths of these lines are the base of the *recurrence quantification analysis* (RQA) [52, 53, 54].

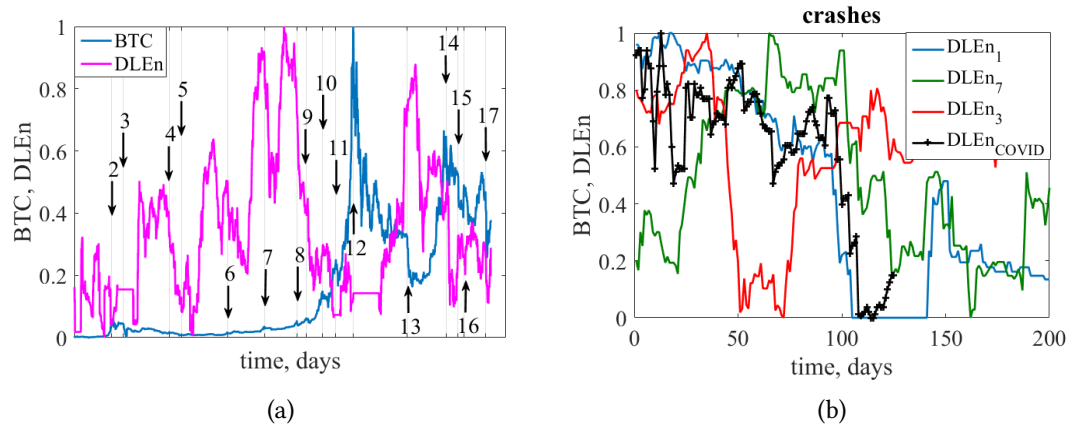
A large number of different approaches have been developed to obtain as much as possible information about the nature of the studied phase space. An important class of recurrence quantifiers is those that try to capture the level of complexity of a signal. In accordance with this study, the *entropy diagonal line histogram* (DLEn) is of the greatest interest which uses the Shannon entropy of the distribution of diagonal lines  $P(l)$  to determine the complexity of the diagonal structures within the recurrence plot. One of the most know quantitative indicators of the recurrence analysis can be defined as:

$$DLEn = - \sum_{l=l_{min}}^{l=l_{max}} p(l) \log p(l)$$

and

$$p(l) = \frac{P(l)}{\sum_{l=l_{min}}^N P(l)},$$

where  $p(l)$  captures the probability that the diagonal line has the exactly length  $l$ , and  $DLEn$  reflects the complexity of deterministic structure in the system. Further calculations were provided and presented in figure 4 for both Bitcoin time series 4a and its several local crashes 4b.



**Figure 4:** DLEn dynamics along with the entire time series of Bitcoin (a). The dynamics of DLEn for the local crashes (b) in accordance with the table 1

However, as follows from the analysis of the entropy indicators, the results may differ for different data preparation. Thus, further in the paper, we take into account two types of en-

tropy based on the general Shannon's approach: *recurrence period density entropy* (RPDEn) and *recurrence entropy* (RecEn).

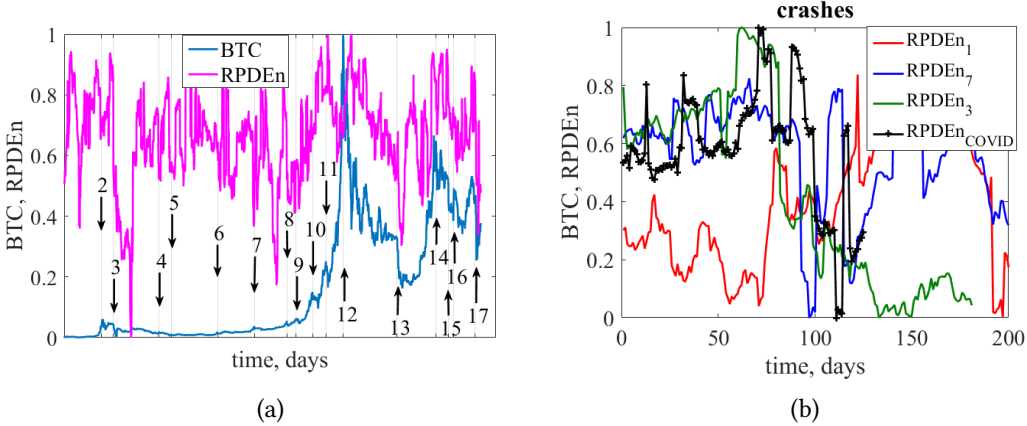
The RPDEn is the quantitative measure of the recurrence analysis that is useful for characterizing the periodicity or absolutely random processes in the time series. It is useful for quantifying the degree of repetitiveness [55, 56]. Considering embedded data points  $\vec{X}(i)$  and suitable threshold  $\epsilon$  in  $d_E$ -dimensional space, we are following forward in time until it has left this ball of radius  $\epsilon$ . Subsequently, the time  $j$  at which the trajectory first returns to this ball is recorded, and the time difference  $T$  of these two states is recorded. The procedure is repeated for all states of the embedded vector, forming a histogram of recurrence times  $R(T)$ . The histogram is then normalized to give the *recurrence time probability density*:

$$P(T_i) = \frac{R(T_i)}{\sum_{i=1}^{T_{max}} R(T_i)},$$

where  $T_{max} = \max \{T_i\}$ . The normalized entropy of the obtained density can be defined as:

$$RPDEn = \frac{-\sum_{i=1}^{T_{max}} P(T_i) \ln P(T_i)}{\ln T_{max}}. \quad (6)$$

In fact, based on the length of the sequences of neighboring points in the phase space: the more points are neighborhoods, the lower the value of the entropy according to equation (6). The comparing of RPDEn and the Bitcoin's critical states can be seen in figure 5.



**Figure 5:** RPDEn dynamics along with the entire time series of Bitcoin (a). The dynamics of RPDEn for the local crashes (b) in accordance with the table 1

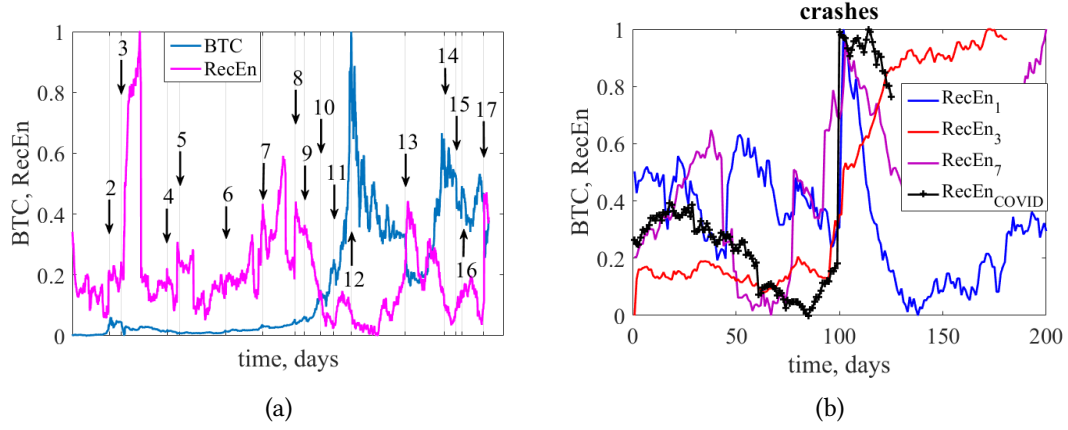
However, recent articles [57, 58] present a slightly different technique for calculating recurrent entropy using a novel way to extract information from the recurrence matrix. To properly define it, we need to define the microstates  $F(\epsilon)$  for the RP that are associated with features of the dynamics of the time series. Selecting the appropriate metric and using the Heaviside function, we evaluate the matrices of dimension  $N \times N$  that are sampled from the RP. The total number of microstates for a given  $N$  is  $N_{ms} = 2^{N^2}$ . The microstates are populated by  $\bar{N}$  random

samples obtained from the recurrence matrix such that  $\bar{N} = \sum_{i=1}^{N_{ms}} n_i$ , where  $n_i$  is the number of times that a microstate  $i$  is observed.

The probability of occurrence of the related microstate  $i$  can be obtained as  $P_i = n_i \cdot (\bar{N})$ . The RecEn of the RP associated with the probability distribution of the corresponding microstates is given by the following equation:

$$RecEn = \sum_{i=1}^{N_{ms}} P_i \ln P_i. \quad (7)$$

In figure 6 we can see the performance of RecEn accordingly to the described above method.



**Figure 6:** RecEn dynamics along with the entire time series of Bitcoin (a). The dynamics of RecEn for the local crashes (b) in accordance with the table 1

## 5. Conclusions

Definitely, the situation with coronavirus is of paramount importance and is of significant danger. From the literature overview, we have understood that the peak and ultimate duration of the outbreak is going to be undetermined for a long time. Yesterday's, today's, and tomorrow's events associated with this pandemic will not disappear without a trace, but will also affect the fate of both individuals and the States in which we live in the long term.

In order to give reliable, powerful, and simple indicators- precursors that are able to minimize further losses as a result of critical changes, we addressed the theory of complexity and the methods of nonlinear dynamics that can identify special trajectories in the complex dynamics and classify them.

The obtained quantitative methods were applied to classified crashes of the Bitcoin market, where it was seen that these indicators can be used to protect yourself from the upcoming critical change. Due to the nature of some crashes and critical events, at any moment there does not exist such a predictive model that could foreshadow them in advance. Thus, the most understandable crashes were selected along with the last critical fall that was caused by COVID-19 due to the desire to understand whether this crash was predictable or not. We

tested 6 measures of complexity on the familiar and past crises, along with the latest ongoing crash, to demonstrate their effectiveness. According to our result, it can be said that general Shannon entropy might be a robust indicator of crashes and critical events not only in the stock market but also in the cryptocurrency market, whereas 5 other measures can serve as efficient harbingers of these events.

Apparently, the impact of the pandemic was reflected in the cryptocurrency market, and therefore, the beginning of the subsequent crisis could be predicted using the appropriate indicators of the theory of complexity. In our further studies, we are going to continue exploring and analyzing other methods from the theory of complexity, and particularly, make research on the fields of artificial intelligence, machine learning, and deep learning [59, 60, 61, 62, 63, 64, 65].

## References

- [1] J. Lonski, Coronavirus may be a black swan like no other, [https://www.moodyanalytics.com/-/media/article/2020/weekly-market-outlook-coronavirus-may-be-black-swan-like-no-other.pdf?source=news\\_body\\_link](https://www.moodyanalytics.com/-/media/article/2020/weekly-market-outlook-coronavirus-may-be-black-swan-like-no-other.pdf?source=news_body_link), 2020.
- [2] T. Fetzter, L. Hensel, J. Hermle, C. Roth, Coronavirus perceptions and economic anxiety, *The Review of Economics and Statistics* (2020) 1–36. doi:10.1162/rest\_a\_00946.
- [3] M. Feldkircher, F. Huber, M. Pfarrhofer, Measuring the effectiveness of us monetary policy during the covid-19 recession, 2020. arXiv:2007.15419.
- [4] R. Cerqueti, V. Ficcadenti, Anxiety for the pandemic and trust in financial markets, 2020. arXiv:2008.01649.
- [5] P. F. Procacci, C. E. Phelan, T. Aste, Market structure dynamics during covid-19 outbreak, 2020. arXiv:2003.10922.
- [6] A. A. Toda, Susceptible-infected-recovered (sir) dynamics of covid-19 and economic impact, 2020. arXiv:2003.11221.
- [7] R. Anderson, J. Heesterbeek, D. Klinkenberg, T. Hollingsworth, Comment how will country-based mitigation measures influence the course of the covid-19 epidemic?, *The Lancet* 395 (2020) 921–1010. doi:10.1016/S0140-6736(20)30567-5.
- [8] B. M. Pavlyshenko, Regression approach for modeling covid-19 spread and its impact on stock market, 2020. arXiv:2004.01489.
- [9] M. Costola, M. Iacopini, C. R. M. A. Santagiustina, Public concern and the financial markets during the covid-19 outbreak, 2020. arXiv:2005.06796.
- [10] K. Arias-Calluari, F. Alonso-Marroquin, M. Nattagh-Najafi, M. Harré, Methods for forecasting the effect of exogenous risk on stock markets, 2020. arXiv:2005.03969.
- [11] M. Garcin, J. Klein, S. Laaribi, Estimation of time-varying kernel densities and chronology of the impact of COVID-19 on financial markets, Working Papers hal-02901988, HAL, 2020. URL: <https://ideas.repec.org/p/hal/wpaper/hal-02901988.html>.
- [12] A. Ammy-Driss, M. Garcin, Efficiency of the financial markets during the COVID-19 crisis: time-varying parameters of fractional stable dynamics, Working Papers hal-02903655, HAL, 2020. URL: <https://ideas.repec.org/p/hal/wpaper/hal-02903655.html>.

- [13] A. F. Colladon, S. Grassi, F. Ravazzolo, F. Violante, Forecasting financial markets with semantic network analysis in the covid-19 crisis, 2020. *arXiv:2009.04975*.
- [14] N. Courtois, M. Grajek, R. Naik, Optimizing sha256 in bitcoin mining, *Communications in Computer and Information Science* 448 (2014) 131–144. doi:10.1007/978-3-662-44893-9\_12.
- [15] L. Kristoufek, Grandpa, grandpa, tell me the one about bitcoin being a safe haven: New evidence from the covid-19 pandemic, *Frontiers in Physics* 8 (2020) 296. URL: <https://www.frontiersin.org/article/10.3389/fphy.2020.00296>. doi:10.3389/fphy.2020.00296.
- [16] D. Broomhead, G. P. King, Extracting qualitative dynamics from experimental data, *Physica D: Nonlinear Phenomena* 20 (1986) 217 – 236. URL: <http://www.sciencedirect.com/science/article/pii/016727898690031X>. doi:[https://doi.org/10.1016/0167-2789\(86\)90031-X](https://doi.org/10.1016/0167-2789(86)90031-X).
- [17] M. Rajkovic, Extracting meaningful information from financial data, *Physica A: Statistical Mechanics and its Applications* 287 (2000) 383–395. doi:10.1016/S0378-4371(00)00377-0.
- [18] V. Ponomarenko, M. Prokhorov, Extracting information masked by the chaotic signal of a time-delay system, *Physical review. E, Statistical, nonlinear, and soft matter physics* 66 (2002) 026215. doi:10.1103/PhysRevE.66.026215.
- [19] M. Henry, G. Judge, Permutation Entropy and Information Recovery in Nonlinear Dynamic Economic Time Series, *Econometrics* 7 (2019) 1–16. URL: <https://ideas.repec.org/a/gam/jecnm/v7y2019i1p10-d213039.html>.
- [20] H. Sigaki, M. Perc, H. Valentin Ribeiro, Clustering patterns in efficiency and the coming-of-age of the cryptocurrency market, *Scientific Reports* 9 (2019) 1440. doi:10.1038/s41598-018-37773-3.
- [21] S. Pincus, R. E. Kalman, Irregularity, volatility, risk, and financial market time series, *Proceedings of the National Academy of Sciences* 101 (2004) 13709–13714. URL: <https://www.pnas.org/content/101/38/13709>. doi:10.1073/pnas.0405168101.
- [22] A. Delgado-Bonal, Quantifying the randomness of the stock markets, *Scientific Reports* 9 (2019). doi:10.1038/s41598-019-49320-9.
- [23] S. Çalik, K. A. Metin, Entropy approach for volatility of wind energy, *Thermal Science* 23 (2019) 1863–1874.
- [24] D. T. Pele, M. Mazurencu, Using high-frequency entropy to forecast bitcoin’s daily value at risk, *Entropy* 21 (2019) 102. doi:10.3390/e21020102.
- [25] A. Belinskyi, V. Soloviev, S. Semerikov, V. Solovieva, Detecting stock crashes using levy distribution, in: A. Kiv, S. Semerikov, V. Soloviev, L. Kibalnyk, H. Danylchuk, A. Matviychuk (Eds.), *Proceedings of the 8th. International Conference on Monitoring, Modeling & Management of Emergent Economy*, volume 2422 of *POPL ’79*, CEUR Workshop Proceedings, Odessa, Ukraine, 2019, pp. 226–236. doi:10.1145/567752.567774.
- [26] A. Bielinskyi, S. Semerikov, V. Solovieva, V. Soloviev, Levy’s stable distribution for stock crash detecting, *SHS Web Conf.* 65 (2019) 06006. URL: <https://doi.org/10.1051/shsconf/20196506006>. doi:10.1051/shsconf/20196506006.
- [27] V. Derbentsev, S. Semerikov, O. Serdyuk, V. Solovieva, V. Soloviev, Recurrence based entropies for sustainability indices, *E3S Web Conf.* 166 (2020) 13031. URL: <https://doi.org/10.1051/e3sconf/202016613031>. doi:10.1051/e3sconf/202016613031.

- [28] V. Soloviev, A. Belinskiy, Complex Systems Theory and Crashes of Cryptocurrency Market: 14th International Conference, ICTERI 2018, Kyiv, Ukraine, May 14-17, 2018, Revised Selected Papers, 2019, pp. 276–297. doi:10.1007/978-3-030-13929-2\_14.
- [29] V. Soloviev, A. Belinskiy, Methods of nonlinear dynamics and the construction of cryptocurrency crisis phenomena precursors, in: V. Ermolayev, M. C. Suárez-Figueroa, V. Yakovyna, V. Kharchenko, V. Kobets, H. Kravtsov, V. Peschanenko, Y. Prytula, M. Nikitchenko, A. Spivakovsky (Eds.), Proceedings of the 13th International Conference on ICT in Education, Research and Industrial Applications. Integration, Harmonization and Knowledge Transfer, volume 2104, CEUR Workshop Proceedings, Kyiv, Ukraine, 2018, pp. 116–127.
- [30] V. Soloviev, O. Serdiuk, Quantum econophysical precursors of cryptocurrency crashes, Cherkasy University Bulletin: Applied Mathematics. Informatics (2020) 3–16. doi:10.31651/2076-5886-2019-1-3-16.
- [31] V. Soloviev, A. Bielinskiy, V. Solovieva, Entropy analysis of crisis phenomena for djia index, in: V. Ermolayev, F. Mallet, V. Yakovyna, V. Kharchenko, V. Kobets, A. Kornilowicz, H. Kravtsov, M. Nikitchenko, S. Semerikov, A. Spivakovsky (Eds.), Proceedings of the 15th International Conference on ICT in Education, Research and Industrial Applications. Integration, Harmonization and Knowledge Transfer, volume 2393, CEUR Workshop Proceedings, Kherson, Ukraine, 2019, pp. 434–449.
- [32] V. Soloviev, O. Serdiuk, S. Semerikov, O. Kohut-Ferens, Recurrence entropy and financial crashes, in: Proceedings of the 2019 7th International Conference on Modeling, Development and Strategic Management of Economic System (MDSMES 2019), Atlantis Press, 2019/10, pp. 385–388. URL: <https://doi.org/10.2991/mdsmes-19.2019.73>. doi:https://doi.org/10.2991/mdsmes-19.2019.73.
- [33] V. Soloviev, S. Semerikov, V. Solovieva, Lempel-ziv complexity and crises of cryptocurrency market, in: Proceedings of the III International Scientific Congress Society of Ambient Intelligence 2020 (ISC-SAI 2020), Atlantis Press, 2020, pp. 299–306. URL: <https://doi.org/10.2991/aebmr.k.200318.037>. doi:https://doi.org/10.2991/aebmr.k.200318.037.
- [34] V. Soloviev, A. Bielinskiy, O. Serdyuk, V. Solovieva, S. Semerikov, Lyapunov exponents as indicators of the stock market crashes, in: O. Sokolov, G. Zholtkevych, V. Yakovyna, Y. Tarasich, V. Kharchenko, V. Kobets, O. Burov, S. Semerikov, H. Kravtsov (Eds.), Proceedings of the 16th International Conference on ICT in Education, Research and Industrial Applications. Integration, Harmonization and Knowledge Transfer. Volume II: Workshops, Kharkiv, Ukraine, October 06-10, 2020, volume 2732 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2020, pp. 455–470. URL: <http://ceur-ws.org/Vol-2732/20200455.pdf>.
- [35] V. Soloviev, V. Solovieva, A. Tuliakova, Visibility graphs and precursors of stock crashes, *Neuro-Fuzzy Modeling Techniques in Economics* (2019) 3–29. doi:10.33111/nfmte.2019.003.
- [36] A. Bielinskiy, S. Semerikov, O. Serdyuk, V. Solovieva, V. Soloviev, L. Pichl, Econophysics of sustainability indices, *CEUR Workshop Proceedings* 2713 (2020) 372–392.
- [37] R. Clausius, *The Mechanical Theory of Heat: With its Applications to the Steam-Engine and to the Physical Properties of Bodies*, J. Van Voorst, 1867.
- [38] J. W. Gibbs, *Elementary Principles in Statistical Mechanics: Developed with Especial Ref-*



- erence to the Rational Foundation of Thermodynamics, Cambridge Library Collection - Mathematics, Cambridge University Press, 2010. doi:10.1017/CBO9780511686948.
- [39] L. Boltzmann, Weitere studien über das wärmeleichgewicht unter gasmolekülen, Wiener Berichte 66 (1872) 275–370.
- [40] C. Chakrabarti, I. Chakrabarty, Boltzmann-shannon entropy: Generalization and application, Modern Physics Letters B 20 (2006) 1471–1479. URL: <https://doi.org/10.1142/S0217984906011529>. doi:10.1142/S0217984906011529.
- [41] C. E. Shannon, A mathematical theory of communication, The Bell System Technical Journal 27 (1948) 379–423. doi:10.1002/j.1538-7305.1948.tb01338.x.
- [42] T. Constantino, Introduction to Nonextensive Statistical Mechanics: Approaching a Complex World, 1st ed., Springer-Verlag New York, 2009.
- [43] R. Sole, S. Valverde, Information Theory of Complex Networks: On Evolution and Architectural Constraints, volume 207, 2004, pp. 189–207. doi:10.1007/978-3-540-44485-5\_9.
- [44] J. S. Richman, J. R. Moorman, Physiological time-series analysis using approximate entropy and sample entropy, American Journal of Physiology-Heart and Circulatory Physiology 278 (2000) H2039–H2049. URL: <https://doi.org/10.1152/ajpheart.2000.278.6.H2039>. doi:10.1152/ajpheart.2000.278.6.H2039. arXiv:<https://doi.org/10.1152/ajpheart.2000.278.6.H2039>, PMID: 10843903.
- [45] S. M. Pincus, Approximate entropy as a measure of system complexity, Proceedings of the National Academy of Sciences 88 (1991) 2297–2301. URL: <https://www.pnas.org/content/88/6/2297>. doi:10.1073/pnas.88.6.2297. arXiv:<https://www.pnas.org/content/88/6/2297.full.pdf>.
- [46] C. Bandt, B. Pompe, Permutation entropy: a natural complexity measure for time series., Physical review letters 88 17 (2002) 174102.
- [47] J. Amigó, Permutation Complexity in Dynamical Systems: Ordinal Patterns, Permutation Entropy and All That (Springer Series in Synergetics), 2010th ed., Springer, Reading, MA., 2010.
- [48] M. Zanin, L. Zunino, O. Rosso, D. Papo, Permutation entropy and its main biomedical and econophysics applications: A review, Entropy 14 (2012) 1553. doi:10.3390/e14081553.
- [49] H. Kantz, T. Schreiber, Nonlinear Time Series Analysis, 2 ed., Cambridge University Press, 2003. doi:10.1017/CBO9780511755798.
- [50] R. Gu, Multiscale shannon entropy and its application in the stock market, Physica A: Statistical Mechanics and its Applications 484 (2017) 215 – 224. URL: <http://www.sciencedirect.com/science/article/pii/S0378437117304740>. doi:<https://doi.org/10.1016/j.physa.2017.04.164>.
- [51] F. Takens, Detecting strange attractors in turbulence, volume 898, 1981, p. 366. doi:10.1007/BFb0091924.
- [52] C. L. Webber, J. P. Zbilut, Dynamical assessment of physiological systems and states using recurrence plot strategies, Journal of Applied Physiology 76 (1994) 965–973. URL: <https://doi.org/10.1152/jappl.1994.76.2.965>. doi:10.1152/jappl.1994.76.2.965. arXiv:<https://doi.org/10.1152/jappl.1994.76.2.965>, PMID: 8175612.
- [53] J. P. Zbilut, C. L. Webber, Embeddings and delays as derived from quantifica-

- tion of recurrence plots, *Physics Letters A* 171 (1992) 199 – 203. URL: <http://www.sciencedirect.com/science/article/pii/037596019290426M>. doi:[https://doi.org/10.1016/0375-9601\(92\)90426-M](https://doi.org/10.1016/0375-9601(92)90426-M).
- [54] N. Marwan, N. Wessel, U. Meyerfeld, A. Schirdewan, J. Kurths, Recurrence-plot-based measures of complexity and their application to heart-rate-variability data, *Physical Review E* 66 (2002) 026702. doi:10.1103/PhysRevE.66.026702.
- [55] H. Rabarimanantsoa, L. Achour, C. Letellier, A. Cuvelier, J.-F. Muir, Recurrence plots and shannon entropy for a dynamical analysis of asynchronisms in noninvasive mechanical ventilation, *Chaos (Woodbury, N.Y.)* 17 (2007) 013115. doi:10.1063/1.2435307.
- [56] M. Little, P. Mcsharry, S. Roberts, D. Costello, I. Moroz, Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection, *Biomedical engineering online* 6 (2007) 23. doi:10.1186/1475-925X-6-23.
- [57] G. Corso, T. L. Prado, G. Z. dos S. Lima, S. R. Lopes, A novel entropy recurrence quantification analysis, 2017. arXiv:1707.00944.
- [58] S. Lopes, T. Prado, G. Corso, G. dos S. Lima, J. Kurths, Parameter-free quantification of stochastic and chaotic signals, *Chaos, Solitons & Fractals* 133 (2020) 109616. URL: <http://www.sciencedirect.com/science/article/pii/S0960077920300151>. doi:<https://doi.org/10.1016/j.chaos.2020.109616>.
- [59] H. Jang, J. Lee, An empirical study on modeling and prediction of bitcoin prices with bayesian neural networks based on blockchain information, *IEEE Access* 6 (2018) 5427–5437.
- [60] L. Alessandretti, A. ElBahrawy, L. M. Aiello, A. Baronchelli, Anticipating cryptocurrency prices using machine learning, *Complexity* 2018 (2018) 1–16. URL: <http://dx.doi.org/10.1155/2018/8983590>. doi:10.1155/2018/8983590.
- [61] L. Zheng, H. He, Share price prediction of aerospace relevant companies with recurrent neural networks based on pca, 2020. arXiv:2008.11788.
- [62] N. Gandal, H. Halaburda, Can we predict the winner in a market with network effects? competition in cryptocurrency market, *SSRN Electronic Journal* (2016). doi:10.2139/ssrn.2832836.
- [63] T. Guo, A. Bifet, N. Antulov-Fantulin, Bitcoin volatility forecasting with a glimpse into buy and sell orders, 2018 IEEE International Conference on Data Mining (ICDM) (2018). URL: <http://dx.doi.org/10.1109/ICDM.2018.00123>. doi:10.1109/icdm.2018.00123.
- [64] T. Guo, N. Antulov-Fantulin, Predicting short-term bitcoin price fluctuations from buy and sell orders, *ArXiv abs/1802.04065* (2018).
- [65] A. Bielinskyi, I. Khvostina, A. Mamanazarov, A. Matviychuk, S. Semerikov, O. Serdyuk, V. Solovieva, V. Soloviev, Predictors of oil shocks. Econophysical approach in environmental science, *IOP Conference Series: Earth and Environmental Science* 628 (2021). doi:10.1088/1755-1315/628/1/012019, 8th International Scientific Conference on Sustainability in Energy and Environmental Science, ISCSEES 2020 ; Conference Date: 21 October 2020 Through 22 October 2020.

# Classification of motor vibration with machine learning methods and simulating the vibration using statistical models

Christoph Kammerer<sup>a</sup>, Micha Küstner<sup>a,b</sup>, Michael Gaust<sup>b</sup>, Pascal Starke<sup>b</sup>, Roman Radtke<sup>a</sup> and Alexander Jesser<sup>a</sup>

<sup>a</sup>University of Applied Sciences Heilbronn, Max-Planck-Str. 39, 74081 Heilbronn, Germany

<sup>b</sup>CeraCon GmbH, Talstraße 2, 97990 Weikersheim, Germany

## Abstract

Reducing costs is an important part in today's business. Therefore manufacturers try to reduce unnecessary work processes and storage costs. Machine maintenance is a big, complex, regular process. In addition, the spare parts required for this must be kept in stock until a machine fails. In order to avoid a production breakdown in the event of an unexpected failure, more and more manufacturers rely on predictive maintenance for their machines. This enables more precise planning of necessary maintenance and repair work, as well as a precise ordering of the spare parts required for this. A large amount of past as well as current information is required to create such a predictive forecast about machines. With the classification of motors based on vibration, this paper deals with the implementation of predictive maintenance for thermal systems. There is an overview of suitable sensors and data processing methods, as well as various classification algorithms. In the end, the best sensor-algorithm combinations are shown.

## Keywords

Predictive Maintenance, Industry 4.0, Internet of Things, Big Data, Industrial Internet

## 1. Introduction

The topic of predictive maintenance (PMA) is becoming more and more important for industrial plants and is the key topic in mechanical engineering from the Industry 4.0 aspect [1]. PMA is defined as condition-based maintenance which is carried out on the basis of a wear or service life forecast [2].

PMA uses methods that allow for individual maintenance intervals of an industrial plant to be determined and the maintenance process to be initiated automatically. As part of a R&D cooperation project between CeraCon GmbH and the Heilbronn University of Applied Sciences, a thermal system is to be set up under automation and a PMA strategy is to be implemented, which should then be adaptable to other industrial plants<sup>1</sup>. Due to the complexity of industrial


---

CS&SE@SW 2020: 3rd Workshop for Young Scientists in Computer Science & Software Engineering, November 27, 2020, Kryvyi Rih, Ukraine

✉ christoph.kammerer@hs-heilbronn.de (C. Kammerer)



© 2020 Copyright for this paper by its authors.  
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

<sup>1</sup>Sponsored by ZIM ZF4644801.

plants, an intelligent solution is required in order to be able to offer individual maintenance strategies depending on the state of the plant. For this reason, the project uses machine learning (ML) methods.

The essential steps of an intelligent PMA strategy are the digital acquisition of (sensor) data, their evaluation, the analysis of the acquired data and the prediction of probable events.

First, possible component defect combinations (CDC) of the industrial plant were analyzed using standard technical risk analysis methods (FMEA, risk graph, fault tree analysis) [3]. CDC is the assignment of a wear component of the industrial system to a potentially occurring defect. Depending on the number of possible defects, a component can therefore have several CDCs. Each CDC was assigned an potential detection measure, e.g. physical vibration measurement or electrical current measurement. Suitable sensors were selected for the analyzed detection measures and analyzed with regard to the PMA strategy. CDC's with the same detection methods were combined and measurement data recorded with the respective sensors.

The core of this work is the evaluation of a combination of detection measures for data processing methods and ML algorithms. The optimal combination of these is a prerequisite for an efficient PMA strategy that can be used for the respective industrial plant.

### **1.1. State of the Art**

A study by Bearingpoint [4] shows that PMA implementations capture 76% of the relevant data using suitable sensors, although only 59% of the process, measurement and machine data are evaluated in a targeted manner. There are three basic approaches to implementing a PMA strategy [5].

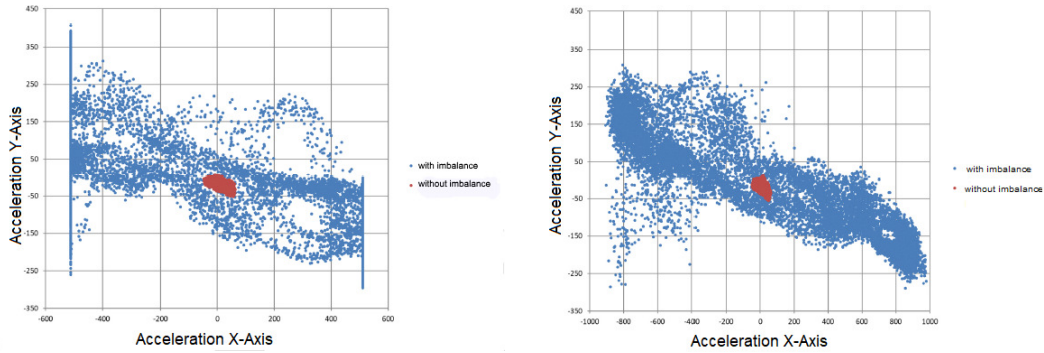
A basic approach is to use the already implemented sensors of the plant for process monitoring. This passive method is particularly suitable for systems that are already in operation. Another passive approach is to introduce dedicated sensors into the system. The additional sensors are introduced to monitor defined wear components and to detect potential defects. In the third approach, a test signal is actively fed into the system. The degree of wear of the components to be monitored can be deduced from the feedback. An example of this is Time Domain Reflectometry (TDR) [5].

## **2. Data Collection**

### **2.1. Sensor Resolution**

When buying industrial sensors, you often have to commit to a sensor resolution. This requires that you have a basic understanding of what accelerations occur on the component. For this purpose, the effects were previously considered in an experiment when an accelerometer with an insufficient resolution is used. In this case, the sensor generates vibrations that exceed the sensor resolution. A CDC of the fan motor is that the fan wheel has an imbalance. This fault situation was simulated by attaching an unbalance to the fan blade.

The result of this simulation is shown in figure 1 (a). There are shown the measured acceleration values in x- and y- axis of an acceleration sensor with a maximum resolution of  $\pm 2G$ . The red values show the vibrations of the motor without an imbalance and the blue values show



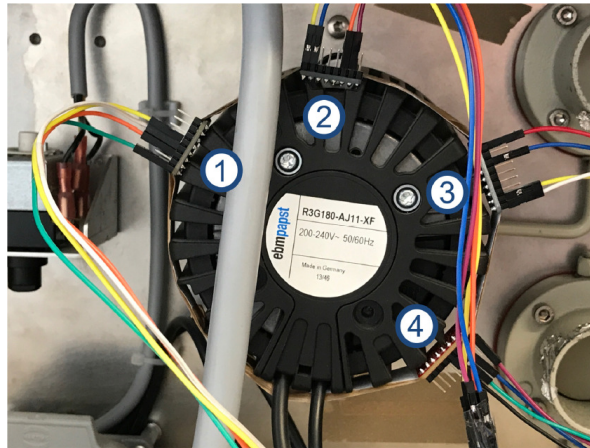
**Figure 1:** Comparison of an unbalanced fan with a resolution of  $\pm 2G$  (a) and a resolution of  $\pm 4G$  (b).

the vibrations which occurs with an imbalance. It can be clearly seen that the vibrations on the motor increased due to the imbalance. It can also be seen that vibrations that go beyond the set sensor resolution of  $\pm 2G$  were not recorded correctly. They are in line with the maximum acceleration of  $\pm 2G$ . The measured values that did not exceed the maximum resolution were not affected by this. The experiment shows that the CDC "imbalance" can cause very strong vibrations. The vibrations are so strong that they exceed a sensor resolution of  $\pm 2G$ . If a sensor is used that can only record values up to a resolution/acceleration of  $\pm 2G$ , these are recorded incorrectly. The values that exceed the maximum resolution are then incorrectly saved in the data record [6]. To prevent such problems, it is important to see how large the vibrations can be. The sensor resolution should have at least this value with a safety buffer. In figure 1 (b), instead of the resolution of  $\pm 2G$ , the double resolution of  $\pm 4G$  was chosen for the same motor level. In the picture you can see that no "lines" have formed and therefore the vibrations were not greater than the sensor resolution. The resolution of  $\pm 4G$  is therefore much more suitable than the resolution of  $\pm 2G$ . The experiment has shown that a correctly selected sensor resolution is a prerequisite for obtaining meaningful results. If the vibrations are greater than the resolution of the sensor, the incorrectly stored measured values cannot be classified correctly [6].

## 2.2. Sensors and Test Set-Up

The requirement for a condition-based PMA is a structured data collection of sensor values. The following sensors were used to obtain status data:

- Three-axis acceleration sensors (Accelerometer):
  - LIS 3DH, MMA 8451, ADXL 343, ADXL 345
- Three-axis acceleration sensors with three-axis yaw rate sensor (gyroskope)
  - MPU 60.50
- Three-axis magnetic field sensor (magnetometer)
  - MLX 90393



**Figure 2:** Measurement setup with the three-axis acceleration sensors LIS3DH, MMA 8451, ADXL 345 and ADXL 343.

**Table 1**

A Section of the Data measured with the MPU 60.50 Gyroscope.

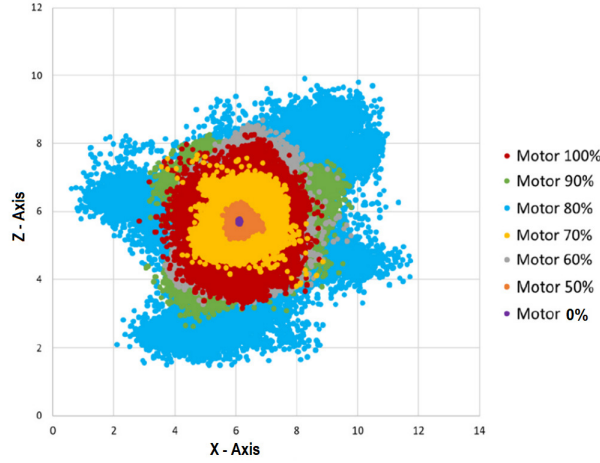
60.50 AccelX	60.50 AccelY	60.50 AccelZ	60.50 GyroX	60.50 GyroY	60.50 GyroZ	Target
416	6088	60146	2709	62478	65187	Motor 000
426	6026	60148	2733	62469	65185	Motor 000
404	6110	60146	2727	62478	65192	Motor 000
470	6046	60140	2720	62486	65188	Motor 000

- Multi sensors with three-axis acceleration, three-axis yaw rate and three-axis magnetic field measurement
  - MPU 92.65, BNO 055, GY 250, GY 521

These recorded the acceleration, the rotation rate and the surrounding magnetic field of the fan motor R3G180-AJ11-XF from ebmpapst Muldingen GmbH & Co. KG used in the thermal system. Figure 2 shows the measurement set-up with the selected three-axis acceleration sensors [6].

The fan motor was operated at fixed speeds, which were divided into 7 classes. This classification was based on the specific values 0%, 50%, 60%, 70%, 80%, 90% and 100% of the maximum engine speed. During the operation of the fan motor the vibration of the crankcase was sensed and recorded by the sensors. More than 980,000 structured sensor data sets per measurement series and sensor type were recorded. A total of more than 2.6 million data sets have thus been recorded for all sensor types. A section of a full data set is shown in Table 1.

An example of a recorded data set is shown in figure 3. It shows the measured acceleration from the housing vibration in the spatial x- and z- orientation. The individual classes are highlighted in color to make a distinction possible. Due to the highest spatial coverage, it can be seen that the measurement results for class 80% can be assigned to the resonance range of the fan motor, since the acceleration values in the x- and z- alignment are at their maximum values



**Figure 3:** Measurement results from the acceleration sensor MMA 8451.

here.

### 3. Data Conditioning

In order to be able to better differentiate the individual classes, it is in some cases advantageous if the data records are processed before classification. The methods used for data conditioning are presented here:

One possibility to process the data sets consists of the differencing and absolute value formation of subsequent values according to equation 1.

$$X_i = |X_i - X_{i+1}| \quad (1)$$

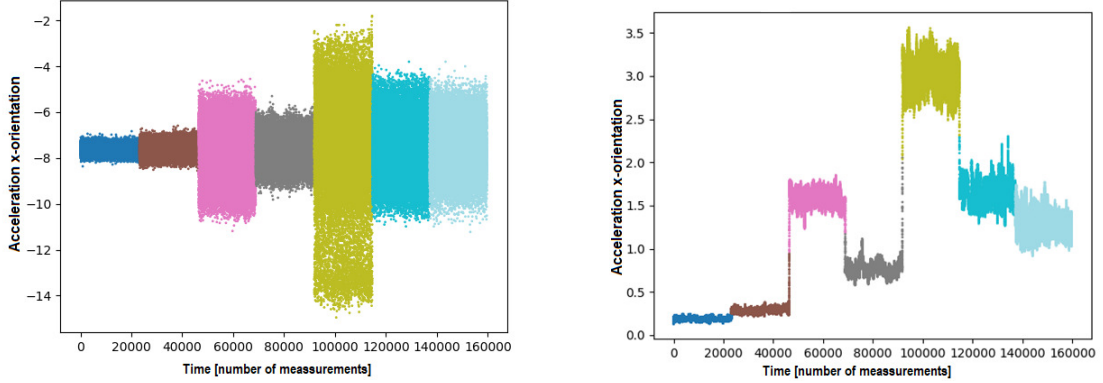
$X_i$  and  $X_{i+1}$  are the successive sensor values. Another processing method is the integration of the data according to equation 2. Here the area under two successive values  $X_i$  and  $X_{i+1}$  is calculated.

$$X_i = \begin{cases} X_i + 0.5 \cdot (X_{i+1} - X_i) & \text{if } X_i < X_{i+1} \\ X_i - 0.5 \cdot (X_i - X_{i+1}) & \text{if } X_i > X_{i+1} \\ X_i & \text{if } X_i = X_{i+1} \end{cases} \quad (2)$$

In both the processing methods, an additional smoothing can be carried out by calculating the moving average according to equation 3.

$$X_i = \frac{1}{G} \cdot \sum_{i-g}^{i+g} X_i \quad (3)$$

The parameter  $G$  specifies the degree of smoothing. The parameter  $g$  is the difference between the indices between the instantaneous value  $X_i$  and the maximum value  $X_{g \pm i}$  specified by



**Figure 4:** Comparison of raw data (a) and data prepared by differencing (b).

the degree of smoothing. Thus  $g$  depends on the degree of smoothing  $G$  and can be determined according to equation 4.

$$g = \frac{G - 1}{2} \quad (4)$$

With the degree of smoothing  $G$ , first optimizations regarding the classification of the measured sensor data can be carried out [7]. Figure 4 shows the effect of processing by means of differencing compared to the unprocessed raw data. The coloring in the pictures illustrates the different class assignments.

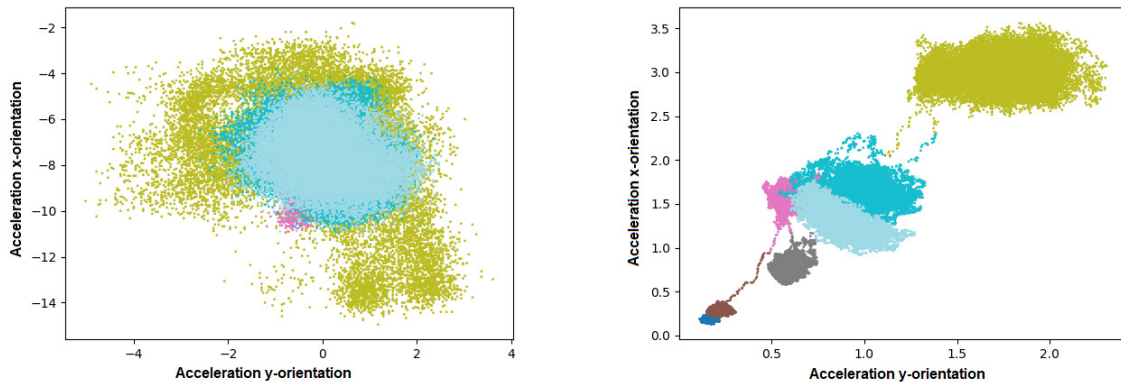
Figure 4 (a) shows the acquired raw data of an acceleration sensor in the x-orientation. It can be seen that a delimitation regarding the classes is not clear. For example, the acceleration in the direction of the x-axis at  $-8m/s^2$  is not unique and can in principle be assigned to any class. In Figure 4 (b) the classes are more delimited after the differencing and smoothing and thus a class assignment is clearer. For example, the value 0.7 can be clearly assigned to the class shown in gray.

Figure 5 (a) shows the raw data from the measurements in x- and y-orientation. A classification is clearly not possible due to the overlapping point clouds. Figure 5 (b), on the other hand, shows the data prepared after the differencing. It can be seen that the point clouds are now clearly distinguishable, making visual and algorithmic class assignment easier.

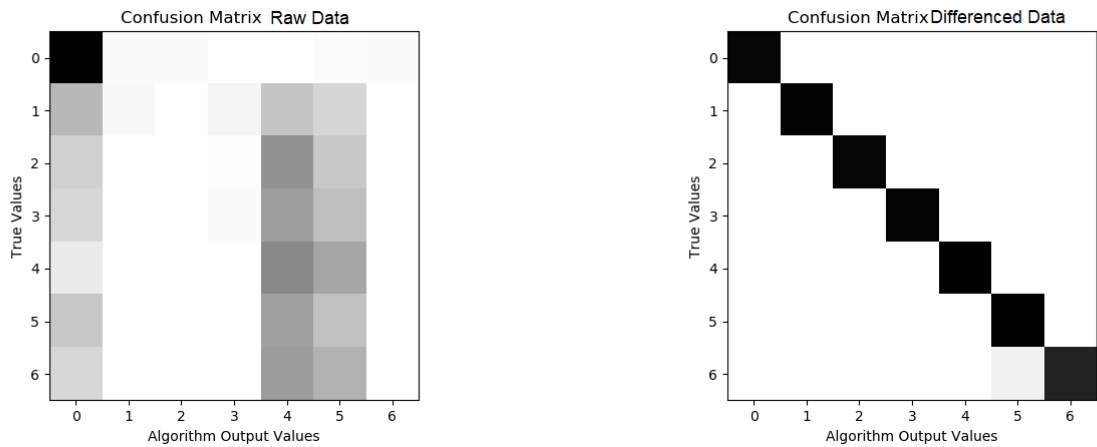
## 4. Evaluation

The evaluation of the ML algorithms with regard to the respective sensors and the data processing was divided into a training and a test phase. In the training phase, the data records were divided evenly by feeding every tenth data value of the respective training method to the ML algorithm. As a result, the respective ML algorithm was trained with 10% of the data. The complete data set was then evaluated in the test phase. Several ML algorithms were considered for the recorded data sets. These included decision trees [8, 9], the gradient boost method





**Figure 5:** Comparison of raw data (a) and data prepared by differencing (b) in two axes.



**Figure 6:** Comparison of the prediction results of the focus cluster algorithm with raw data (a) and data prepared by differencing (b).

[8, 10], a focus cluster algorithm [11] and artificial neural networks (ANN) [12]. The investigations revealed that ANNs are less suitable for data sets with low attribute numbers due to the long duration in the training phase. Therefore, only the decision trees, the gradient boost method and the focus cluster algorithm were used for the further experiments [7].

Figure 6 shows two confusion matrices [13] for the focus cluster algorithm, which show the distribution between the actual class and the class determined by the algorithm. The numbers on the axes correspond to the seven defined classes in which the data records have been categorized. The darker an area, the more often the ML algorithm has assigned data records to a class. A correct assignment is obtained if the assigned class corresponds to the actual class. Ideally, you would get a black diagonal from top left to bottom right.

Figure 6 (b) shows the result for the data sets prepared after differencing and smoothing. It

can be seen that the majority of the data records were assigned to the actual classes, the hit rate here was over 98%. On the other hand, it can be seen in Figure 6 (a) that a significantly lower hit rate has been achieved for the unprepared data sets.

## 5. Using Statistical Methods for Prediction

The preceding test were performed by measuring the vibration of a stand-alone motor on a workbench, which was not build into a working machine. Therefore this data cannot be used to make a prediction for maintenance, but the feasibility of categorizing a motor by its vibration and magnetic field was studied. To get a better picture of the real working conditions of such a motor a larger data set was collected by mounting three multifunction sensors (GY 521, BNO 055, MPU 92.65) on a thermal system which is used in day to day operations. These data sets were then used to build a statistical model based on auto regression and moving average (ARMA [14], [15]) of the vibration. The statistical models were created for every sensor orientation separately to get an optimal result for each time series. As the metric to compare the different models was chosen the maximum relative deviation (MRD) according to equation 5.

$$MRD = \max_i \left| \frac{X_i - Y_i}{X_i} \right| \quad (5)$$

In equation 5  $X_i$  are the measured sensor values used to build the model and  $Y_i$  are the values predicted by the model at this time-step.

## 6. Results

To compare the results of the ML algorithms for the respective sensors, a matrix with the relevant properties was created for each combination of ML algorithm and sensor:

- Classification accuracy (performance) of the algorithms
- Computing time for the training and testing phase of the algorithms
- Smoothing factor G

Table 2 shows the comparison of the processing methods examined using the combination of the multifunction sensor GY 521 and the gradient boost method.

It can be seen that the highest performance is achieved when using the raw data. In the training and test phases the integrated and the differenced data are slightly faster. A comparison was made for each sensor and ML algorithm combination. The best performing data processing method was then selected for each combination. The comparison tables of the sensors which have achieved the best results of the sensor types examined are listed in Tables 3 to 5. These were the ADXL 345 (accelerometer), the MPU 60.50 (gyroscope) and the GY 521 (accelerometer, gyroscope and magnetic field). The processing method with the highest performance for the respective algorithm is shown for each of the sensors.

**Table 2**

Comparison of the data preparation methods with the combination of the multifunction sensor GY521 and the gradient boost method.

Sensor GY 521 with Gradient Boost method	Performance (%)		Time training (s)		Time testing (s)	
	$G = 0$	$G = 99$	$G = 0$	$G = 99$	$G = 0$	$G = 99$
Raw data	98.51	83.97	4.85	6.88	1.11	0.9
Integration	82.84	85.15	5.45	3.82	1.07	0.66
Differencing	79.76	85.1	5.04	3.83	1.5	0.61

**Table 3**

Comparison of the ML algorithms on the combination ADXL 345 (accelerometer) and the best performing data processing method.

ADXL 345	Decision tree	Gradient boost method	Focus cluster algorithm
Conditioning method	Differencing, $G = 99$	Differencing, $G = 99$	Differencing, $G = 99$
Performance	93.79%	96.6%	98.98%
Computing time training	1.78s	5.88s	0.32s
Computing time testing	31.99s	1.44s	34.88s

**Table 4**

Comparison of the ML algorithms on the combination MPU 60.50 (gyroskope) and the best performing data processing method.

MPU 60.50	Decision tree	Gradient boost method	Focus cluster algorithm
Conditioning method	Raw data, $G = 0$	Raw data, $G = 0$	Differencing, $G = 99$
Performance	91.46%	95.24%	90.89%
Computing time training	0.09s	5.54s	0.91s
Computing time testing	0.28s	1.46s	31.99s

The result of the examinations according to Table 3 was that all acceleration sensors achieved the greatest performance with smoothing ( $G = 99$ ) and data prepared by differencing. The focus cluster algorithm achieved the highest performance.

The result according to Table 4 is that the highest performance was achieved with unprocessed and unsmoothed data with the gyroscopes. The gradient boost process achieved the highest performance.

In the case of the multifunction sensors with acceleration, magnetic field sensors and gyroscope, it can be seen from Table 5 that the highest performance was achieved with smoothing ( $G = 99$ ) and differenced data using the cluster cluster algorithm.

Table 6 shows the lowest MRDs for each separate sensor orientation with their number of auto regressive (p) and moving average (q) terms.

These results show a maximum deviation up 28.3% with the gyroscope values in z orientation. The lowest deviation was reached with acceleration in z orientation and the magnetic field in x orientation with only 10.1% and 10.6% respectively. The difference in accuracy of the

**Table 5**

Comparison of the ML algorithms on the combination GY 521 and the best performing data processing method.

GY 521	Decision tree	Gradient boost method	Focus cluster algorithm
Conditioning method	Raw data, $G = 0$	Raw data, $G = 0$	Differencing, $G = 99$
Performance	95.7%	98.51%	99.89%
Computing time training	0.09s	4.85s	1.27s
Computing time testing	0.28s	1.11s	37.63s

**Table 6**

Lowest MRD for each Sensor Orientation with their corresponding p and q Numbers.

Sensor Orientation	p	q	MRD (%)
Acc X	5	5	20.3
Acc Y	3	3	24.4
Acc Z	2	5	10.1
Gyro X	7	5	22.4
Gyro Y	4	3	24.7
Gyro Z	7	6	28.3
Mag X	4	5	10.6
Mag Y	3	5	22.6
Mag Z	7	7	28.2

models is rather high and therefore an ARMA model can only be used to model two of the nine time series measured. For the remaining seven there should be used other means of modeling.

## 7. Conclusion

It has been found that the processing of the raw data in the form of smoothing and differencing in combination with the focus cluster algorithm gave the best results for acceleration sensors. The gyroscopes examined showed that the unprocessed raw data without smoothing in combination with the gradient boost method achieved the highest classifiability. The multisensors examined gave the best results when using the focus cluster algorithm in combination with smoothed and differenced data. In addition it was found, that an ARMA model could be used to predict the acceleration in z orientation and the magnetic field in x orientation.

## 8. Outlook

Based on these results, the combination of detection measure, data processing method and ML algorithm can in the next step be used for a PMA strategy. For a complete PMA, further detection measures have to be examined. For that purpose, this procedure is continued with further sensor types in order to find an optimal combination for all necessary detection measures. In the future, a prediction model is to be developed on the basis of these results, with

which predictions can be made about the degree of wear of system components of a thermal system under automation. Formal aging and error models of the respective system components must also be created in order to map the aging process of components. These models can then be used to make probabilistic statements about the failure probabilities of the individual assemblies. Such models could be based on Dynamic Bayesian Networks (DBN) [16], auto regression and moving average (ARMA) [17] or, as the focus cluster algorithm has yielded such an high performance, a multi dimensional focus trajectory. In addition to that, the statistical models used to predict the motor vibration in day to day operations could be extended to auto regression, integrated, moving average to get a better result for all sensor orientations.

## References

- [1] S. Feldmann, R. Lässig, O. Herweg, H. Rauen, P.-M. Synek, Predictive Maintenance: Servicing tomorrow ? and where we are really at today, Technical Report, Roland Berger GmbH, München, 2017. URL: <https://www.vdma.org/documents/105806/17180011/VDMA+Predictive+Maintenance+englisch.pdf>.
- [2] Din en 13306:2018-02: Instandhaltung - begriffe der instandhaltung; dreisprachige fassung en 13306:2017, 2018. URL: <https://www.beuth.de/en/standard/din-en-13306/270274780>. doi:10.31030/2641990.
- [3] B. Ebert, Prozessoptimierung bei Industrie 4.0 durch Risikoanalysen: Gefährdungen erkennen und minimieren, Springer Vieweg, Berlin, Heidelberg, 2018. doi:10.1007/978-3-662-55729-7.
- [4] F. Duscheck, R. Blameuser, S. Gehrman, Consulting, Solutions and Ventures | Bearing-Point, 2018. URL: <https://www.bearingpoint.com>.
- [5] H. M. Hashemian, W. C. Bean, State-of-the-art predictive maintenance techniques\*, IEEE Transactions on Instrumentation and Measurement 60 (2011) 3480–3492. doi:10.1109/TIM.2009.2036347.
- [6] M. Küstner, Auswahl und bewertung eines sensor-konzepts für die implementierung einer predictive maintenance zur temperaturbehandlung unter automation, 2019.
- [7] M. Gaust, Entwurf und implementierung von maschinellen lernalgorithmen zur klassifikation von maschinendaten für eine vorausschauende wartung von industriethermosystemen, 2019.
- [8] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, Édouard Duchesnay, Scikit-learn: Machine learning in python, Journal of Machine Learning Research 12 (2011) 2825–2830. URL: <http://jmlr.org/papers/v12/pedregosa11a.html>.
- [9] T. Thomas, A. P. Vijayaraghavan, S. Emmanuel, Applications of Decision Trees, Springer Singapore, Singapore, 2020, pp. 157–184. URL: [https://doi.org/10.1007/978-981-15-1706-8\\_9](https://doi.org/10.1007/978-981-15-1706-8_9). doi:10.1007/978-981-15-1706-8\_9.
- [10] A. Binding, N. Dykeman, S. Pang, Machine learning predictive maintenance on data in the wild, in: 2019 IEEE 5th World Forum on Internet of Things (WF-IoT), 2019, pp. 507–512. doi:10.1109/WF-IoT.2019.8767312.

- [11] A. Géron, *Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques for Building Intelligent Systems*, 1st ed., O'Reilly Media, Boston, 2017.
- [12] T. Rashid, *Make Your Own Neural Network*, 1st ed., CreateSpace Independent Publishing Platform, North Charleston, SC, USA, 2016.
- [13] M. Füllsack, A (partially) interactive introduction to Systems Sciences, 2013. URL: <http://systems-sciences.uni-graz.at/etextbook/bigdata/confusionmatrix.html>.
- [14] C. Lin, Y. Hsieh, F. Cheng, H. Huang, M. Adnan, Time series prediction algorithm for intelligent predictive maintenance, *IEEE Robotics and Automation Letters* 4 (2019) 2807–2814. doi:10.1109/LRA.2019.2918684.
- [15] H. Akaike, Maximum likelihood identification of gaussian autoregressive moving average models, *Biometrika* 60 (1973) 255–265. URL: <http://www.jstor.org/stable/2334537>.
- [16] P. Dagum, A. Galper, E. Horvitz, Dynamic network models for forecasting, in: D. Dubois, M. P. Wellman, B. D'Ambrosio, P. Smets (Eds.), *Uncertainty in Artificial Intelligence*, Morgan Kaufmann, 1992, pp. 41–48. URL: <https://www.sciencedirect.com/science/article/pii/B9781483282879500104>. doi:<https://doi.org/10.1016/B978-1-4832-8287-9.50010-4>.
- [17] J. Wang, Y. Liang, Y. Zheng, R. X. Gao, F. Zhang, An integrated fault diagnosis and prognosis approach for predictive maintenance of wind turbine bearing with limited samples, *Renewable Energy* 145 (2020) 642–650. URL: <https://www.sciencedirect.com/science/article/pii/S0960148119309371>. doi:<https://doi.org/10.1016/j.renene.2019.06.103>.

# Features of the development of an automated educational and control complex for checking the quality of students

Dmitriy Bukreiev, Pavlo Chorny, Evgeniy Kupchak and Andrey Sender

*Bogdan Khmelnytsky Melitopol State Pedagogical University, 20 Getmanska Str., Melitopol, 72300, Ukraine*

## Abstract

The paper reveals the problematic features of developing an automated system for teaching students. The authors emphasize the need to study the specialized needs of the system before the development process and conduct an in-depth analysis of the needs of the modern educational process in order to identify the needs of the future product. In this paper authors develop a model of knowledge representation in the system and reveal the features of the mathematical apparatus for calculating and automating the processes of obtaining and evaluating students knowledge.

## Keywords

development, educational, control complex, automated educational system

## 1. Introduction

Innovations in higher education are determined by state policy, one of the important priorities of which is informatization of all spheres of activity. The practice of modernizing higher education shows that with the help of traditional means of training and control, it is increasingly difficult to solve the problems of training modern personnel. In accordance with it, as well as with the requirements of the state standard of higher professional education in higher educational institutions, the role of training using ICT has significantly increased.

Many scientists have made a significant contribution to the theory and practice of using ICT as learning tools: Oleksandr Yu. Burov [1], Valerii Yu. Bykov [2, 3, 4], Svitlana M. Hryshchenko [5, 6], Arnold E. Kiv [7, 8, 9, 10, 11], Vladyslav S. Kruglyk [12, 13], Oksana M. Markova [14, 15], Mykhailo M. Mintii [16], Yevhenii O. Modlo [17, 18, 19, 20, 21], Pavlo P. Nechypurenko [22], Natalya V. Rashevskaya [23], Kateryna P. Osadcha [24], Viacheslav V. Osadchyi [25, 26, 27, 28, 29], Serhiy O. Semerikov [30, 31, 32, 33], Kateryna I. Slovak [34], Andrii M. Striuk [35], Svitlana V. Symonenko [36, 37], Viktoriia V. Tkachuk [38, 39], Tetiana A. Vakaliuk [40], Nataliia V. Valko [41, 42], Yuliia V. Yechkalo [43] and others. Methods of using the distance learning tools and the problems of blended learning were analyzed by Olga V. Bondarenko [44], Pavlo I. Fedoruk

*CS&SE@SW 2020: 3rd Workshop for Young Scientists in Computer Science & Software Engineering, November 27, 2020, Kryvyi Rih, Ukraine*

✉ [dessnight@gmail.com](mailto:dessnight@gmail.com) (D. Bukreiev); [pavlochornuj@gmail.com](mailto:pavlochornuj@gmail.com) (P. Chorny); [kupchak@mdpu.org.ua](mailto:kupchak@mdpu.org.ua) (E. Kupchak); [sender@mdpu.org.ua](mailto:sender@mdpu.org.ua) (A. Sender)

🆔 0000-0002-5150-153X (D. Bukreiev); 0000-0002-8239-0211 (P. Chorny); 0000-0003-1659-1450 (E. Kupchak); 0000-0002-8741-5625 (A. Sender)

© 2020 Copyright for this paper by its authors.  
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).



CEUR Workshop Proceedings (CEUR-WS.org)

[45], Nadiia O. Holiver [46], Volodymyr M. Kukharenko [47], Larysa M. Petrenko [48], Svitlana V. Shokaliuk [49], Mariya P. Shyshkina [50], Oleg M. Spirin [51], Myroslav J. Syvyi [52] and others. In their researches ways and ways of increase of efficiency of training with use of modern computer engineering, problems of computerization of natural sciences are considered, classifications of pedagogical software are offered. In the works [53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65] addressed the use of computer technology in teaching general professional discipline, the use of this technique for modeling processes and phenomena, compiling tests, solving problems and conducting quantitative calculations, developing software for self-control and standardized control of knowledge, computerization of pedagogical experiment.

Our theoretical analysis of the literature has shown that the use of new information technologies allows the student to better understand the nature of the course being studied, actively participate in the process of mastering its content, and independently monitor the acquired knowledge, skills and abilities. For their part, the teacher can adapt software tools in accordance with specific learning goals, quickly and objectively identify the level of development of educational material by students.

However, as for the use of new information technologies in courses of general professional disciplines, not all theoretical and scientific-methodological issues have been solved here. It can be stated that there is a mismatch between the needs of higher education institutions in the use of computer-based information learning technologies and insufficient development of the corresponding software packages.

As noted in the work of Viacheslav V. Osadchyi and Iryna V. Krashenninnik, nowadays one of the main deterrents to the improvement of professional training in such conditions is lower, compared to the standard term of study for a bachelor's degree (4 years) to 2-3 years of study with subsequent transition to professional training. However, during the two years of study it is impossible to provide the necessary time for full practical training [66], so the creation of an individual trajectory of the student allows to intensify the workload and reduces the time to obtain general competencies.

The question arises as to which of the modern approaches should be relied on to eliminate this inconsistency and thus improve the quality of teaching of general professional disciplines. One of the most relevant and promising in this regard is the competency approach, aimed at forming in future professionals key and a number of actual professional competencies during the shortened period of study.

## **2. Features of the educational process**

The pedagogical process is always two-way and two-way and the analysis of educational and pedagogical goals, which sets the "object" of pedagogical influence, is no less (if not more) important than the analysis of the teacher's goals and the development of his pedagogical creativity. As noted by A.K. Markov, the level of development of goal-setting processes is the most important indicator of the formation of educational activities. We are talking about the quality of goals (novelty, non-standard, flexibility, sustainability), their validity and realism, the ability to overcome obstacles to achieving them. In the course of training there is a change of interaction of the educational purposes and motives, internal motives of educational activity,



in particular, cognitive and professional are formed. The most important factor in intensifying this process was the possibility of independent setting of educational goals. Classifications of goals act as a special measure of their commonality:

- global, general and private goals,
- attitude to educational structures that are responsible for their formulation and achievement,
- substructures of personality, the development of which they focus,
- goal description language.

Information technology is inextricably linked to forecasting because it has common goals. They are focused on achieving specific results, but require more rigor and responsibility in the design for direct application, which uses heuristic techniques, while forecasting allows extrapolation.

Important in our opinion in the study is the work of Viacheslav V. Osadchyi and Kateryna P. Osadcha, which identifies current trends in education informatization, among which scientists highlight: the introduction and dissemination of e-learning, services and tools for teachers, students, leaders and parents to cooperate with all the above participants in the learning process, the development of cloud technologies [67]. The authors emphasize that these tasks are leading in the scientific space. However, before determining the quality of the graduate obtained at the end of training, it is necessary to determine the factors that have the greatest impact on the learning process of students [68]. And as further noted in the study [69], the requirements of employers for engineering and technical knowledge and skills are constantly becoming more complicated. This is due to the accelerated evolution of technical skills, the emergence of new engineering professions and the penetration of technology into all spheres of human life.

The interaction of information technology for the teaching of general professional disciplines gives reason to believe that in the design of pedagogical technologies, new sources of forecasting are possible.

Pedagogical technologies with the use of information technology in a military university are associated with modeling, which, as a means of representation and transformation of an object that does not yet exist in reality, allows:

- lose, compare and evaluate standard pedagogical technologies,
- to simulate real processes of future pedagogical activity,
- accept the result of one of the alternative solutions to pedagogical problems.

The didactic model of teaching a general professional discipline as a variant of pedagogical technology should allow, if necessary, to make it possible to identify individual essential aspects of future activities, subjecting them to a more thorough logical analysis. Such modeling allows you to operate with unidentified objects and determine their stable properties, which is

extremely important in the military education system. The concept of pedagogical technologies, aimed at identifying their definition, structure, classification and justification of choice, is based on the analysis of theories of various sciences and related concepts.

In the process of professional training and scientific search, knowledge, skills and abilities are formed, there is a comprehensive intellectual development of students' personality and world-view, mastering the chosen specialty. All this happens through the main sphere – the mental activity of students: sensation, perception, representation, comprehension, memorization and other mental processes. As a result of mental functioning, all this is analyzed, synthesized through the harmonious action of higher nervous activity of a person. On the basis of this activity, a learning system is created that combines the content and form of scientific knowledge, establishing connections and relationships between the subjects studied and the phenomena of the objective world.

Trying to classify educational tasks by their "cognitive composition", we can observe that all types of tasks are grouped into five groups:

- tasks for reproduction of knowledge,
- tasks that require simple mental operations,
- tasks that require complex mental operations,
- tasks that require productive thinking for their solution,
- tasks for productive thinking with the generation on its basis of written or oral expression.

The process of professionally-oriented training was usually based on an inductive basis, from partial to general, by consistently performing a number of partial functions inherent in a military specialist. At the same time, each of these functions was studied in isolation from each other without any relationship. Therefore, the process of acquiring professional skills necessary for mastering complex equipment and the process of its operation was delayed.

So, building professional training on an inductive basis does not justify itself today. It seems most appropriate to build the learning process on a deductive basis from the general to the particular, so that professional skills and abilities of students are formed on the basis of general knowledge. This approach is used today to justify a new system and content of vocational training that corresponds to the modern development of vocational education. It requires the use of Information Technology. The new content of training using information technologies should ultimately lead to a change in Information Technologies of training that allow the development of expert training systems for assessing knowledge, skills and abilities. Such expert systems should be based on the principles of the theory of step-by-step formation of mental actions and skills.

Among these principles we highlight the following:

- transition to the planning of the educational process in accordance with the level of knowledge acquisition in general discipline,

- introduction into the educational process of a quantitative measure of the degree of completion of the learning process in the form of the coefficient of assimilation,
- expert-educational system of assessment of knowledge, skills, abilities should be created taking into account the two above-mentioned principles.

Creation of expert-educational, expert systems on an estimation of quality of mastering of knowledge and completeness of process of training provides first of all the account of the basic principles:

- changing the role and function of the teacher, turning him into a specialist “consultant”, which adds a new responsibility in his teaching activities,
- abandonment of the current method of training and the transition to individual training, especially important for self-training,
- transfer of the center of gravity of the educational process to the independent work of students,
- preparation of educational and methodical complex on the basis of taking into account features of computer technology of training. Each student is provided with full textbooks and various tasks in the discipline,
- abandonment of traditional forms of control and the introduction of individual cumulative index, which sharply increases the role of current, boundary and final control of knowledge, skills and abilities.

If the above principles are strictly followed, we can talk about the possibility of developing and using in the educational process of expert training systems and systems of expert assessment of knowledge, skills and abilities, their content of textbooks and manuals.

This indicates the need to increase the practical orientation of student learning. To implement professional training of students in educational standards and objectives requires the development and implementation of scientifically sound methods that reduce the time of students to master the actions and operations of professional activities without reducing the required quality of education.

The analysis of works of Pavlo I. Fedoruk showed that a significant amount of work is focused on forecasting trends in the development of educational space according to modern needs and, in particular, the development of adaptive testing tools for students. In his works, the scientist emphasizes the need to solve a number of problems of test control of knowledge in modern educational systems. In his works, the author notes that the methods and tools of classical test theory, despite the great potential of this technology do not solve many problems posed by the current level of education. He emphasizes that adaptive tests, in this case, can solve this problem and can be effectively used to solve any problem of optimizing the educational process - assessing the effectiveness of pedagogical innovations and technologies, monitoring [70]. Summing up the analysis, we can say that we can identify a number of positive features of the use of adaptive tests, namely:

- adjustment to individual possibilities of the student,
- increasing the accuracy of assessing the level of knowledge,
- reducing the duration of the test,
- reducing the degree of student fatigue,
- ensuring confidentiality by providing each student with an individual set of test tasks that correspond to his level of knowledge,
- simplification of the procedure for making changes to the task bank.

In the works of Valerii Yu. Bykov emphasized that the further informatization of the educational environment is based on computerized means of labor and requires a fundamentally new information environment [2, 3, 4]. Such systems are a further development of expert systems. They provide a quantitative assessment of system performance on the basis of selected criteria and models of alternative situations.

The development of methods of teaching general discipline of students with the help of information technology should be based on the integration of the provisions of the theory of the gradual formation of mental actions and block-module learning.

### **3. Features of system requirements**

Teaching the general professional discipline aims to lay a theoretical foundation for understanding the essence of the physical and mechanical characteristics of the main structural materials and the technology of their production for their use in human activities. To ensure the consistency of knowledge and skills in the general discipline, practical classes are complex in nature. Individualization is achieved through an individual approach to students during practical, laboratory classes, consultations, current and final control. More hours are allocated for self-training of students. The acquired knowledge and practical skills are the basis for studying all subsequent engineering disciplines, they are the basis for technical design, construction and operation of airfields.

The main part of special courses is the organization and conduct of practical classes that make it possible to understand the features of the production process of a particular discipline. Practical classes are the basis of practical technology, so the process of improving the theoretical level of the content of the general discipline is a very important component. Strengthening the role of theoretical knowledge and conducting classes in accordance with the developed thematic plan is insufficient, and there is a change in the ratio of factual and theoretical material in favor of the latter, which leads to insufficient justification of theories in the minds of students and consolidation of acquired skills and knowledge.

A way out of this situation can currently be found when using information technology of training in the process of teaching the general discipline, which allows you to create a methodically based flow of information, including, in particular, factual material, which in the future can become the basis for the manifestation of systematizing and explaining functions

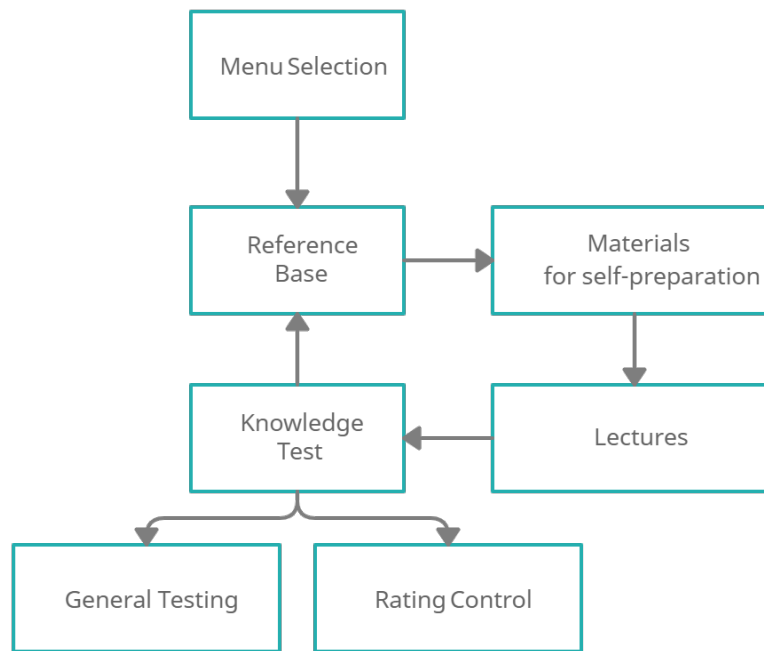
of theoretical knowledge. It is clear that creating an information flow is impossible without using an automated training complex. Information technologies open up an opportunity to better understand the nature of the object itself, actively engage in the process of its cognition, independently changing both its parameters and operating conditions. In this regard, information technologies can not only have a positive impact on the study of the general discipline by students, but, more importantly, on their mental development. The use of information technologies allows you to quickly and objectively identify the level of mastering the material, which is of great importance in the process of teaching a general professional discipline.

The main criterion for assessing the degree of information content of the educational process is the possibility of using training and testing programs in the educational process, which organize active work of students in a dialogue with the teacher. The advantages of using information technologies in the study of general disciplines include:

- improving the quality of teaching as a result of rapid assessment of knowledge,
- study group monitoring,
- operational management of the training session,
- effective and convenient work of the teacher in the group,
- clarity and visual control of knowledge,
- processing and storage of large amounts of information,
- accelerating the assimilation of material,
- acquisition of sustainable skills for practical application in the chosen field,
- manifestation of pedagogical creativity in the methodology and technology of the educational process.

Thus, the use of information technology for classes in general discipline in the system of automated student training does not impose any restrictions on the didactic structure of educational material and the choice of a pedagogical approach, but, on the contrary, contains a stage of creative approach by both teacher and student, organizing a closed interconnected cycle.

Summing up, we can say that, when creating an educational and methodological professionally oriented environment in the teaching of general professional discipline, different approaches are taken into account in developing a didactic model of teaching general professional discipline and include the following components: motivational, meaningful, personal-activity, managerial, evaluative. The selected didactic model reflects the conditions, stages and methods of teaching general professional discipline in the framework of professional training. The development of the model was based on the principles of systematization, scientificity, modeling, autonomy.



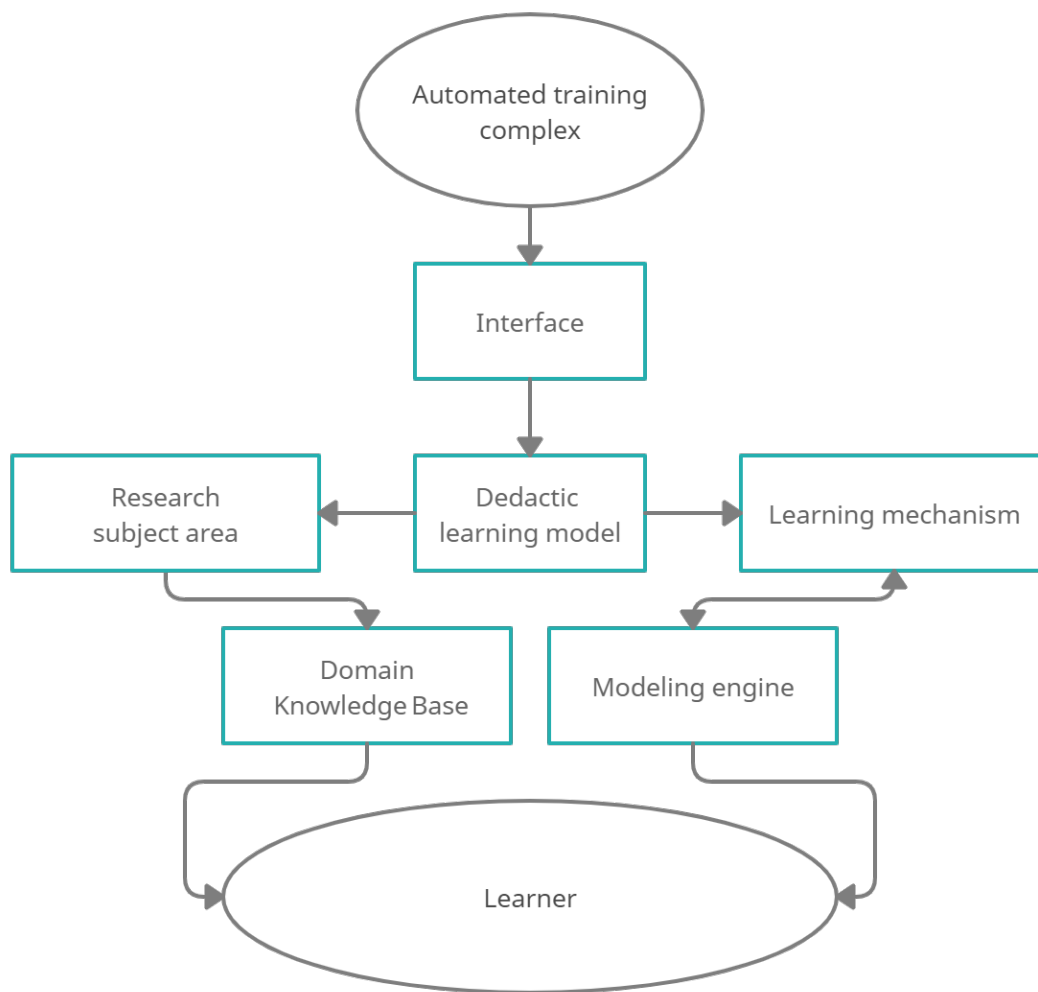
**Figure 1:** Virtual learning environment of the automated training complex

#### 4. Structure of the software complex

On the basis of a tabular list of educational elements of a general professional discipline, didactic learning goals are developed, the achievement of which is controlled by a system of tests for the corresponding levels of assimilation of acquired knowledge. In automated knowledge control, tests are developed using a special methodology that takes into account the low, medium and high level of training of students. All training elements must be covered by testing, the number of tests developed for each training element is unlimited and can be constantly updated, since the training complex has a modular structure. Unlike traditional tests, automated tests involve the content of reference answers in the general discipline. After conducting knowledge testing in a computer classroom, the coefficient of knowledge acquisition by students is calculated, which is the basis of the created expert system for assessing the success of the learning process.

The coefficient of assimilation of knowledge varies from 0 to 1. if the coefficient of assimilation is greater than or equal to 0.7, the learning process is considered completed, specific knowledge is obtained and skills are acquired, and the coefficient of assimilation of knowledge is used in monitoring the rating control when calculating the rating: by semester, by specialty, individually and in a group test, which is extremely important for analyzing the assimilation of the course. The virtual learning environment using an automated training complex has a block structure, which is reflected in the figure 1.

As the automated educational complex of training of general professional discipline corre-



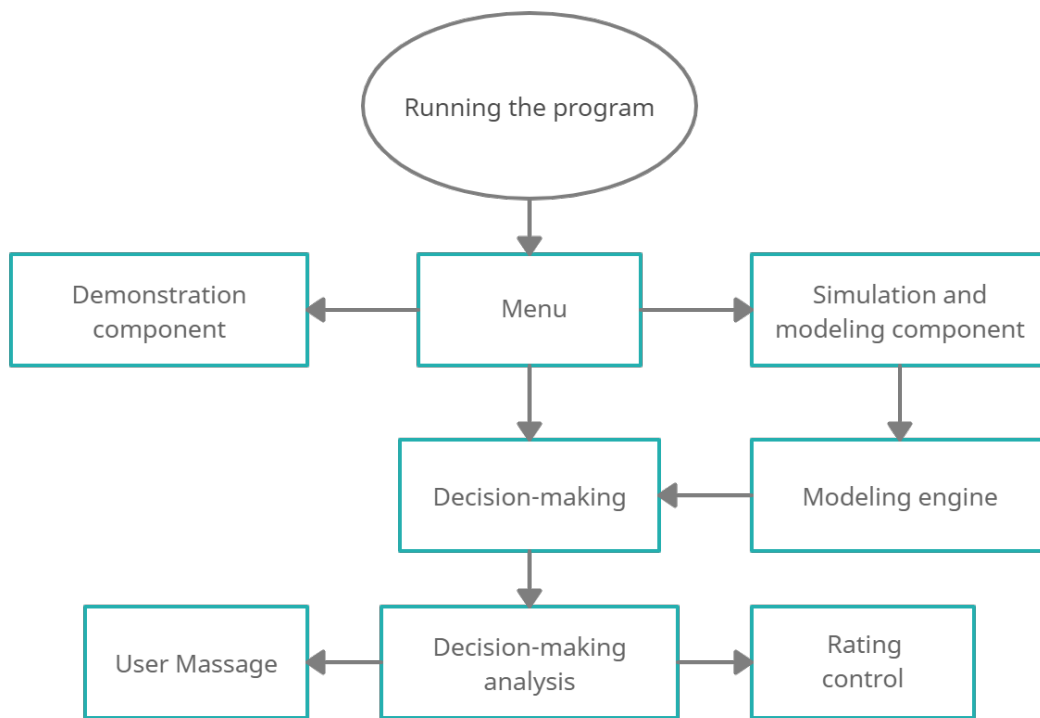
**Figure 2:** Homeostasis structure of knowledge transfer

sponds to the general requirements shown to educational controlling complexes, it is focused on achievement of the purposes set by the teacher and combines traditional and information technologies.

When using an automated training complex, both individual and group training can be used. In the context of learning, the complex has a homeostasis structure of knowledge transfer.

The global purpose of the automated educational complex is to prepare students for full and effective participation in social and professional fields in the information society. At its application there is a change of relations between the teacher and the student. They become a process of joint creativity (figure 2).

The automated training complex is a training, testing and monitoring program compiled according to the didactic model of PC formation when teaching a general professional discipline.



**Figure 3:** The functional structure of the training function of the automated training complex

The program is prepared in the Visual Studio environment with an integrated Shell and special navigation tools for viewing. In the Visual Studio System, there is a special form designer that allows you to prepare windows of the future program in the form of forms. The program is focused on internet technologies that allow you to deploy databases and build multi-level modular software packages with built-in application servers. The program is implemented in the client/server architecture and consists of several modules, each of which is significant and has its own algorithm. Software modules start working from the moment they are launched and do not require any additional configuration.

After registering the user, you need to determine the operating mode on the monitor screen. To conduct self-training or the next lesson, you need to use the menu view of the automated training complex. It contains: training, lecture, testing unit of self-training, monitoring of rating control of knowledge. Depending on the pedagogical task, the appropriate level is selected, and the program works according to the selected branch and a certain algorithm (figure 3).

Automated training complex provides for the presence of three components:

- demonstration, which displays information in accordance with a pre-designed scenario,
- simulation-modeling, which allows the user to control the dynamics of the learning process,



- controls the component that analyzes the user's work.

From a methodological point of view, all parts are in demand. If the educational part of the program is selected, the algorithm of the structure of the educational function of the automated educational complex has the following form (figure 3).

Working with the lecture course, the user works with the electronic textbook, gradually pressing the preview button and first selecting a topic from the list of menu sections on the screen, the thematic plan of the discipline is constantly consistent with the selected learning elements. The student on request can print out any lecture and view the physical properties of the materials from the reference database, learn the goals and objectives of the study, as well as read the list of references.

Simultaneously with the study of theoretical material, you can test your knowledge by choosing the option of testing knowledge. The testing unit for checking the quality of knowledge, skills and abilities reflects the systematic content of the general discipline and provides an objective check of the level of academic achievement. Tests are divided into classes by level:

- tests in practical and laboratory classes – a simple test,
- tests of increased complexity by topic – mixed test,
- residual knowledge tests – a control test for general disciplines.

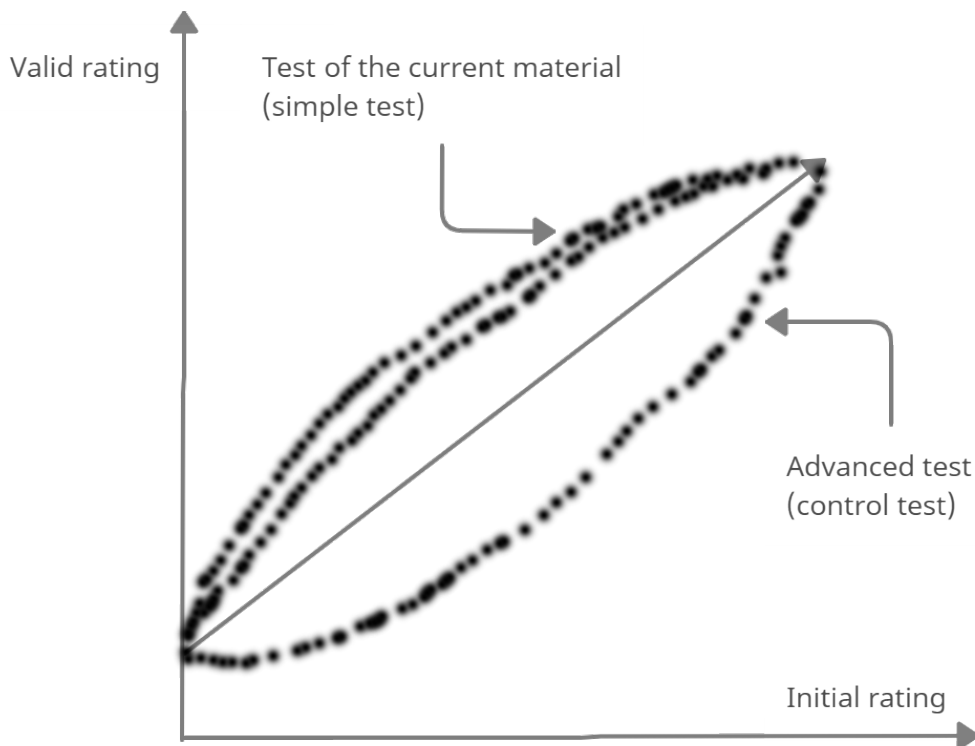
In its form, the test is a system of tasks. Test tasks in the general discipline have a short form, so that each one does not take much time to complete. Work in each test block is individual.

The block of tests based on current material works according to the principle: question – answer based on time characteristics and contains approximate answers, of which you need to choose the correct one. Text appears on the screen, and questions and several possible answers to them are displayed in turn. After answering the questions, a rating is given. The accumulation of individual grades obtained on tests allows monitoring the assimilation of each topic of the general professional discipline both separately for each cadet and for the group as a whole.

When performing tasks on self-preparation for classes, it is sometimes very convenient to use testing on lecture topics of the general professional discipline. Simple tests contain material from practical and laboratory work. In a complex test, a mixed test consists of thirty tests and contains questions on the topic of a general professional discipline, while a control test consists of sixty tests and contains questions to determine in-depth knowledge in a general discipline. It can be used for self-certification and verification of residual knowledge in the general discipline.

The ambiguity of tests developed for teaching a general professional discipline is obvious. Each of them complements each other. The test curve determines the dependence of the test subject's actual rating. Training typical situations are selected by difficulty level only taking into account previous test checks. The curve is the average straight line between the characteristic curve of a light and heavy test (figure 4).

The result of the experimental study is expressed in a test score. Test scores are located on special scales, the score obtained with the help of tests is more accurate and differentiated than



**Figure 4:** Dependences of the actual rating of the tested person on the characteristic curve of the test

traditional. If you look at the distribution of frequencies of actual scores depending on the characteristics of the test after a pedagogical experiment, the abilities of the tested were distributed according to the normal law. Performing tests independently allows cadets to conduct self-preparation for classes and eliminate gaps in the assimilation of lecture material, thereby increasing their professional level.

After studying the general discipline, the total coefficient of mastering of each student is calculated according to the formula:

$$K_a = K_l / i * n,$$

where:  $K_a$  is the total coefficient of assimilation after studying the topic;  $K_l$  - a separate coefficient of assimilation on the  $i$ -th test;  $n$  - the number of topics in the general professional discipline.

At final control of knowledge the received general factor is considered as follows. At  $K_a = 1.0 - 0.9$  the grade "5" is set, at  $K_a = 0.9 - 0.8$  - a mark "4", at  $K_a = 0.8 - 0.7$  - a mark "3" and at  $K_a < 0.7$  - a mark "5".

## 5. Conclusions

The problem features of development of the automated system for training of students were opened in the work. Thus, it was emphasized the need to use a competency-based approach, which in its work is aimed at forming in future professionals key and a number of actual professional competencies during the shortened training period. The pedagogical process is always two-way and two-way and the analysis of educational and pedagogical goals, which sets the "object" of pedagogical influence, is a more important factor than the analysis of the goals of the teacher and the development of his pedagogical creativity. It was emphasized the need to study the specialized needs of the system before the development process and conducted an in-depth analysis of the needs of the modern educational process in order to identify the needs of the future product. During the analysis it was determined that information learning technologies allow the development of expert-educational systems for assessing knowledge, skills and abilities, such expert systems should be based on the principles of the theory of gradual formation of mental actions and skills. Among these principles were the following: the transition to planning the learning process in accordance with the level of knowledge acquisition in general discipline; introduction into the educational process of a quantitative measure of the degree of completion of the learning process in the form of the coefficient of assimilation; expert-educational system of assessment of knowledge, skills, abilities should be created taking into account the two above-mentioned principles. It was emphasized that the development of methods of teaching general discipline of students with the help of information technology should be based on the integration of the provisions of the theory of the gradual formation of mental actions and block-module learning. The developed model of knowledge representation in the system and reveals the features of the mathematical apparatus of calculation and automation of processes of obtaining and evaluating students' knowledge, so the automated educational complex provides for three components: demonstration, which displays information in accordance with pre-developed scenario; simulation-modeling, which allows the user to control the dynamics of the learning process; controls the component, which makes an analysis of the user. The block of tests on the current material works on the principle: the question – the answer taking into account time characteristics and contains approximate answers from which it is necessary to choose the correct one. When performing self-preparation tasks, it is sometimes very convenient to use testing with lecture topics of general professional discipline. Simple tests contain material from practical and laboratory work. In the complex test - a mixed test consists of thirty tests and contains questions on the topic of general professional discipline, and the control test consists of sixty tests and contains questions to determine in-depth knowledge of general discipline. The result of the experimental study is expressed in a test score. Test scores are located on special scales, the score obtained by the tests is more accurate and differentiated than traditional

## References

- [1] O. Burov, A. Kiv, S. Semerikov, A. Striuk, M. Striuk, L. Kolgatina, I. Oliinyk, Aredu 2020 - how augmented reality helps during the coronavirus pandemic, volume 2731, CEUR-WS,

- 2020, pp. 1–46. 3rd International Workshop on Augmented Reality in Education, AREdu 2020 ; Conference Date: 13 May 2020.
- [2] V. Bykov, M. Shyshkina, Emerging technologies for personnel training for IT industry in Ukraine, Institute of Electrical and Electronics Engineers Inc., 2014, pp. 945–949. doi:10.1109/ICL.2014.7017903, 2014 International Conference on Interactive Collaborative Learning, ICL 2014 ; Conference Date: 3 December 2014 Through 6 December 2014.
- [3] V. Bykov, A. Dovgiallo, P. Kommers, Theoretical backgrounds of educational and training technology, International Journal of Continuing Engineering Education and Life-Long Learning 11 (2001) 412–441.
- [4] V. Bykov, A. Gurzhiy, G. Kozlakova, Development of computer education in Ukrainian higher technical schools, A-52, Elsevier Science Publishers B.V., Amsterdam, Netherlands, Hamburg, Ger, 1994. Proceedings of the IFIP 13th World Computer Congress. Part 3 (of 3) ; Conference Date: 28 August 1994 Through 2 September 1994.
- [5] V. Morkun, S. Semerikov, S. Hryshchenko, K. Slovak, Environmental geo-information technologies as a tool of pre-service mining engineer’s training for sustainable development of mining industry, volume 1844, CEUR-WS, 2017, pp. 303–310. 13th International Conference on ICT in Education, Research and Industrial Applications. Integration, Harmonization and Knowledge Transfer, ICTERI 2017 ; Conference Date: 15 May 2017 Through 18 May 2017.
- [6] V. Morkun, S. Semerikov, N. Morkun, S. Hryshchenko, A. Kiv, Defining the structure of environmental competence of future mining engineers: Ict approach, volume 2257, CEUR-WS, 2018, pp. 198–203. 1st International Workshop on Augmented Reality in Education, AREdu 2018 ; Conference Date: 2 October 2018.
- [7] A. Kiv, S. Semerikov, V. Soloviev, A. Striuk, First student workshop on computer science & software engineering, volume 2292, CEUR-WS, 2018, pp. 1–10. 1st Student Workshop on Computer Science and Software Engineering, CS&SE@SW 2018 ; Conference Date: 30 November 2018.
- [8] A. Kiv, V. Soloviev, S. Semerikov, Cte 2018 – how cloud technologies continues to transform education, volume 2433, CEUR-WS, 2019, pp. 1–19. 6th Workshop on Cloud Technologies in Education, CTE 2018 ; Conference Date: 21 December 2018.
- [9] A. Kiv, S. Semerikov, V. Soloviev, A. Striuk, Second student workshop on computer science & software engineering, volume 2546, CEUR-WS, 2019, pp. 1–20. 2nd Student Workshop on Computer Science and Software Engineering, CS&SE@SW 2019 ; Conference Date: 29 November 2019.
- [10] A. Kiv, M. Shyshkina, S. Semerikov, A. Striuk, Y. Yechkalo, Aredu 2019 – how augmented reality transforms to augmented learning, volume 2547, CEUR-WS, 2020, pp. 1–12. 2nd International Workshop on Augmented Reality in Education, AREdu 2019 ; Conference Date: 22 March 2019.
- [11] A. Kiv, M. Shyshkina, S. Semerikov, A. Striuk, M. Striuk, H. Shalatska, Cte 2019-when cloud technologies ruled the education, volume 2643, CEUR-WS, 2020, pp. 1–59. 7th Workshop on Cloud Technologies in Education, CTE 2019 ; Conference Date: 20 December 2019.
- [12] H. Chemerys, V. Osadchyi, K. Osadcha, V. Kruhlyk, Increase of the level of graphic competence future bachelor in computer sciences in the process of studying 3d modeling,

- volume 2393, CEUR-WS, 2019, pp. 17–28. 15th International Conference on ICT in Education, Research and Industrial Applications. Integration, Harmonization and Knowledge Transfer, ICTERI 2019 ; Conference Date: 12 June 2019 Through 15 June 2019.
- [13] V. Kruglyk, V. Osadchyi, Developing competency in programming among future software engineers, *Integration of Education* 23 (2019) 587–606. doi:10.15507/1991-9468.097.023.201904.587-606.
- [14] O. Markova, S. Semerikov, M. Popel, Cocalc as a learning tool for neural network simulation in the special course "foundations of mathematic informatics", volume 2104, CEUR-WS, 2018, pp. 388–403. 14th International Conference on ICT in Education, Research and Industrial Applications. Integration, Harmonization and Knowledge Transfer. Volume II: Workshops, ICTERI 2018 ; Conference Date: 14 May 2018 Through 17 May 2018.
- [15] O. Markova, S. Semerikov, A. Striuk, H. Shalatska, P. Nechypurenko, V. Tron, Implementation of cloud service models in training of future information technology specialists, volume 2433, CEUR-WS, 2019, pp. 499–515. 6th Workshop on Cloud Technologies in Education, CTE 2018 ; Conference Date: 21 December 2018.
- [16] M. Petrova, M. Mintii, S. Semerikov, N. Volkova, Development of adaptive educational software on the topic of "fractional numbers" for students in grade 5, volume 2292, CEUR-WS, 2018, pp. 162–192. 1st Student Workshop on Computer Science and Software Engineering, CS&SE@SW 2018 ; Conference Date: 30 November 2018.
- [17] Y. Modlo, S. Semerikov, Xcos on web as a promising learning tool for bachelor's of electromechanics modeling of technical objects, volume 2168, CEUR-WS, 2017, pp. 34–41. 5th Workshop on Cloud Technologies in Education, CTE 2017 ; Conference Date: 28 April 2017.
- [18] Y. Modlo, S. Semerikov, E. Shmeltzer, Modernization of professional training of electromechanics bachelors: Ict-based competence approach, volume 2257, CEUR-WS, 2018, pp. 148–172. 1st International Workshop on Augmented Reality in Education, AREdu 2018 ; Conference Date: 2 October 2018.
- [19] Y. Modlo, S. Semerikov, P. Nechypurenko, S. Bondarevskiy, O. Bondarevska, S. Tolmachev, The use of mobile internet devices in the formation of ict component of bachelors in electromechanics competency in modeling of technical objects, volume 2433, CEUR-WS, 2019, pp. 413–428. 6th Workshop on Cloud Technologies in Education, CTE 2018 ; Conference Date: 21 December 2018.
- [20] Y. Modlo, S. Semerikov, S. Bondarevskiy, S. Tolmachev, O. Markova, P. Nechypurenko, Methods of using mobile internet devices in the formation of the general scientific component of bachelor in electromechanics competency in modeling of technical objects, volume 2547, CEUR-WS, 2020, pp. 217–240. 2nd International Workshop on Augmented Reality in Education, AREdu 2019 ; Conference Date: 22 March 2019.
- [21] Y. Modlo, S. Semerikov, R. Shajda, S. Tolmachev, O. Markova, P. Nechypurenko, T. Selivanova, Methods of using mobile internet devices in the formation of the general professional component of bachelor in electromechanics competency in modeling of technical objects, volume 2643, CEUR-WS, 2020, pp. 500–534. 7th Workshop on Cloud Technologies in Education, CTE 2019 ; Conference Date: 20 December 2019.
- [22] P. Nechypurenko, S. Semerikov, Vlabembed - the new plugin moodle for the chemistry education, volume 1844, CEUR-WS, 2017, pp. 319–326. 13th International Conference on

- ICT in Education, Research and Industrial Applications. Integration, Harmonization and Knowledge Transfer, ICTERI 2017 ; Conference Date: 15 May 2017 Through 18 May 2017.
- [23] N. Rashevskaya, S. Semerikov, N. Zinonos, V. Tkachuk, M. Shyshkina, Using augmented reality tools in the teaching of twodimensional plane geometry, volume 2731, CEUR-WS, 2020, pp. 79–90. 3rd International Workshop on Augmented Reality in Education, AREdu 2020 ; Conference Date: 13 May 2020.
- [24] K. Osadcha, V. Osadchyi, S. Semerikov, H. Chemerys, A. Chorna, The review of the adaptive learning systems for the formation of individual educational trajectory, volume 2732, CEUR-WS, 2020, pp. 547–558. 16th International Conference on ICT in Education, Research and Industrial Applications. Integration, Harmonization and Knowledge Transfer. Volume II: Workshops, ICTERI 2020 ; Conference Date: 6 October 2020 Through 10 October 2020.
- [25] V. Osadchyi, N. Valko, N. Kushnir, Determining the level of readiness of teachers to implementation of stem-education in ukraine, volume 2393, CEUR-WS, 2019, pp. 144–155. 15th International Conference on ICT in Education, Research and Industrial Applications. Integration, Harmonization and Knowledge Transfer, ICTERI 2019 ; Conference Date: 12 June 2019 Through 15 June 2019.
- [26] V. Osadchyi, H. Varina, E. Prokofiev, I. Serdiuk, S. Shevchenko, Use of ar/vr technologies in the development of future specialists' stress resistance: Experience of steam-laboratory and laboratory of psychophysiological research cooperation, volume 2732, CEUR-WS, 2020, pp. 634–649. 16th International Conference on ICT in Education, Research and Industrial Applications. Integration, Harmonization and Knowledge Transfer. Volume II: Workshops, ICTERI 2020 ; Conference Date: 6 October 2020 Through 10 October 2020.
- [27] V. Osadchyi, I. Krasheninnik, O. Spirin, S. Koniukhov, T. Diuzhykova, Personalized and adaptive ict-enhanced learning: A brief review of research from 2010 to 2019, volume 2732, CEUR-WS, 2020, pp. 559–571. 16th International Conference on ICT in Education, Research and Industrial Applications. Integration, Harmonization and Knowledge Transfer. Volume II: Workshops, ICTERI 2020 ; Conference Date: 6 October 2020 Through 10 October 2020.
- [28] V. Osadchyi, H. Chemerys, K. Osadcha, V. Kruhlyk, S. Koniukhov, A. Kiv, Conceptual model of learning based on the combined capabilities of augmented and virtual reality technologies with adaptive learning systems, volume 2731, CEUR-WS, 2020, pp. 328–340. 3rd International Workshop on Augmented Reality in Education, AREdu 2020 ; Conference Date: 13 May 2020.
- [29] V. Osadchyi, H. Varina, K. Osadcha, O. Prokofieva, O. Kovalova, A. Kiv, Features of implementation of modern ar technologies in the process of psychological and pedagogical support of children with autism spectrum disorders, volume 2731, CEUR-WS, 2020, pp. 263–282. 3rd International Workshop on Augmented Reality in Education, AREdu 2020 ; Conference Date: 13 May 2020.
- [30] S. Semerikov, I. Teplytskyi, Y. Yechkalo, A. Kiv, Computer simulation of neural networks using spreadsheets: The dawn of the age of camelot, volume 2257, CEUR-WS, 2018, pp. 122–147. 1st International Workshop on Augmented Reality in Education, AREdu 2018 ; Conference Date: 2 October 2018.
- [31] S. Semerikov, I. Teplytskyi, Y. Yechkalo, O. Markova, V. Soloviev, A. Kiv, Computer simu-

- lation of neural networks using spreadsheets: Dr. anderson, welcome back, volume 2393, CEUR-WS, 2019, pp. 833–848. 15th International Conference on ICT in Education, Research and Industrial Applications. Integration, Harmonization and Knowledge Transfer, ICTERI 2019 ; Conference Date: 12 June 2019 Through 15 June 2019.
- [32] S. Semerikov, A. Striuk, L. Striuk, M. Striuk, H. Shalatska, Sustainability in software engineering education: A case of general professional competencies, volume 166, EDP Sciences, 2020. doi:10.1051/e3sconf/202016610036, 2020 International Conference on Sustainable Futures: Environmental, Technological, Social and Economic Matters, ICSF 2020 ; Conference Date: 20 May 2020 Through 22 May 2020.
- [33] S. Semerikov, S. Chukharev, S. Sakhno, A. Striuk, V. Osadchyi, V. Solovieva, T. Vakaliuk, P. Nechypurenko, O. Bondarenko, H. Danylchuk, Our sustainable coronavirus future, volume 166, EDP Sciences, 2020. doi:10.1051/e3sconf/202016600001, 2020 International Conference on Sustainable Futures: Environmental, Technological, Social and Economic Matters, ICSF 2020 ; Conference Date: 20 May 2020 Through 22 May 2020.
- [34] S. Semerikov, V. Pototskyi, K. Slovak, S. Hryshchenko, A. Kiv, Automation of the export data from open journal systems to the russian science citation index, volume 2257, CEUR-WS, 2018, pp. 215–226. 1st International Workshop on Augmented Reality in Education, AREdu 2018 ; Conference Date: 2 October 2018.
- [35] A. Striuk, S. Semerikov, The dawn of software engineering education, volume 2546, CEUR-WS, 2019, pp. 35–57. 2nd Student Workshop on Computer Science and Software Engineering, CS&SE@SW 2019 ; Conference Date: 29 November 2019.
- [36] S. Symonenko, V. Osadchyi, S. Sysoieva, K. Osadcha, A. Azaryan, Cloud technologies for enhancing communication of itprofessionals, volume 2643, CEUR-WS, 2020, pp. 225–236. 7th Workshop on Cloud Technologies in Education, CTE 2019 ; Conference Date: 20 December 2019.
- [37] S. Symonenko, N. Zaitseva, V. Osadchyi, K. Osadcha, E. Shmeltser, Virtual reality in foreign language training at higher educational institutions, volume 2547, CEUR-WS, 2020, pp. 37–49. 2nd International Workshop on Augmented Reality in Education, AREdu 2019 ; Conference Date: 22 March 2019.
- [38] V. Tkachuk, Y. Yechkalo, S. Semerikov, M. Kislova, V. Khotskina, Exploring student uses of mobile technologies in university classrooms: Audience response systems and development of multimedia, volume 2732, CEUR-WS, 2020, pp. 1217–1232. 16th International Conference on ICT in Education, Research and Industrial Applications. Integration, Harmonization and Knowledge Transfer. Volume II: Workshops, ICTERI 2020 ; Conference Date: 6 October 2020 Through 10 October 2020.
- [39] V. Tkachuk, S. Semerikov, Y. Yechkalo, S. Khotskina, V. Soloviev, Selection of mobile ict for learning informatics of future professionals in engineering pedagogy, volume 2732, CEUR-WS, 2020, pp. 1058–1068. 16th International Conference on ICT in Education, Research and Industrial Applications. Integration, Harmonization and Knowledge Transfer. Volume II: Workshops, ICTERI 2020 ; Conference Date: 6 October 2020 Through 10 October 2020.
- [40] T. Vakaliuk, V. Kontsedailo, D. Antoniuk, O. Korotun, S. Semerikov, I. Mintii, Using game dev tycoon to develop professional soft competencies for future engineers-programmers, volume 2732, CEUR-WS, 2020, pp. 808–822. 16th International Conference on ICT in Edu-

cation, Research and Industrial Applications. Integration, Harmonization and Knowledge Transfer. Volume II: Workshops, ICTERI 2020 ; Conference Date: 6 October 2020 Through 10 October 2020.

- [41] N. Valko, V. Osadchyi, Education individualization by means of artificial neural networks, volume 166, EDP Sciences, 2020. doi:10.1051/e3sconf/202016610021, 2020 International Conference on Sustainable Futures: Environmental, Technological, Social and Economic Matters, ICSF 2020 ; Conference Date: 20 May 2020 Through 22 May 2020.
- [42] N. Valko, N. Kushnir, V. Osadchyi, Cloud technologies for stem education, volume 2643, CEUR-WS, 2020, pp. 435–447. 7th Workshop on Cloud Technologies in Education, CTE 2019 ; Conference Date: 20 December 2019.
- [43] O. Syrovatskyi, S. Semerikov, Y. Modlo, Y. Yechkalo, S. Zelinska, Augmented reality software design for educational purposes, volume 2292, CEUR-WS, 2018, pp. 193–225. 1st Student Workshop on Computer Science and Software Engineering, CS&SE@SW 2018 ; Conference Date: 30 November 2018.
- [44] O. Bondarenko, S. Mantulenko, A. Pikilnyak, Google Classroom as a tool of support of blended learning for geography students, volume 2257, CEUR-WS, 2018, pp. 182–191. 1st International Workshop on Augmented Reality in Education, AREdu 2018 ; Conference Date: 2 October 2018.
- [45] P. Fedoruk, The technology of creating the learning module in the adaptive system of distance education and knowledge control, Porto, 2005, pp. 419–422. IADIS International Conference on Cognition and Exploratory Learning in Digital Age, CELDA 2005 ; Conference Date: 14 December 2005 Through 16 December 2005.
- [46] N. Holiver, T. Kurbatova, I. Bondar, Blended learning for sustainable education: Moodle-based english for specific purposes teaching at kryvyi rih national university, volume 166, EDP Sciences, 2020. doi:10.1051/e3sconf/202016610006, 2020 International Conference on Sustainable Futures: Environmental, Technological, Social and Economic Matters, ICSF 2020 ; Conference Date: 20 May 2020 Through 22 May 2020.
- [47] V. Kukhareno, T. Oleinik, Open distance learning for teachers, volume 2393, CEUR-WS, 2019, pp. 156–169. 15th International Conference on ICT in Education, Research and Industrial Applications. Integration, Harmonization and Knowledge Transfer, ICTERI 2019 ; Conference Date: 12 June 2019 Through 15 June 2019.
- [48] L. Petrenko, S. Kravets, O. Bazeliuk, L. Maiboroda, I. Muzyka, Analysis of the current state of distance learning in the vocational education and training institutions, volume 166, EDP Sciences, 2020. doi:10.1051/e3sconf/202016610010, 2020 International Conference on Sustainable Futures: Environmental, Technological, Social and Economic Matters, ICSF 2020 ; Conference Date: 20 May 2020 Through 22 May 2020.
- [49] S. Shokaliuk, Y. Bohunenko, I. Lovianova, M. Shyshkina, Technologies of distance learning for programming basics on the principles of integrated development of key competences, volume 2643, CEUR-WS, 2020, pp. 548–562. 7th Workshop on Cloud Technologies in Education, CTE 2019 ; Conference Date: 20 December 2019.
- [50] Y. Krylova-Grek, M. Shyshkina, Blended learning method for improving students' media literacy level, volume 2732, CEUR-WS, 2020, pp. 1272–1285. 16th International Conference on ICT in Education, Research and Industrial Applications. Integration, Harmonization and Knowledge Transfer. Volume II: Workshops, ICTERI 2020 ; Conference Date: 6



October 2020 Through 10 October 2020.

- [51] O. Spirin, V. Oleksiuk, N. Balyk, S. Lytvynova, S. Sydorenko, The blended methodology of learning computer networks: Cloud-based approach, volume 2393, CEUR-WS, 2019, pp. 68–80. 15th International Conference on ICT in Education, Research and Industrial Applications. Integration, Harmonization and Knowledge Transfer, ICTERI 2019 ; Conference Date: 12 June 2019 Through 15 June 2019.
- [52] M. Syvyi, O. Mazbayev, O. Varakuta, N. Panteleeva, O. Bondarenko, Distance learning as innovation technology of school geographical education, volume 2731, CEUR-WS, 2020, pp. 369–382. 3rd International Workshop on Augmented Reality in Education, AREdu 2020 ; Conference Date: 13 May 2020.
- [53] N. Selwyn, The use of computer technology in university teaching and learning: A critical perspective, *Journal of Computer Assisted Learning* 23 (2007) 83–94. doi:10.1111/j.1365-2729.2006.00204.x.
- [54] J. Olson, Putting the computer in its place: A study of teaching with technology, *Journal of Curriculum Studies* 26 (1994) 121–141. doi:10.1080/0022027940260201.
- [55] W. Zhao, Y. Sun, L. Dai, Improving computer basis teaching through mobile communication and cloud computing technology, volume 1, Chengdu, 2010, pp. V1452–V1454. doi:10.1109/ICACTE.2010.5578977, 2010 3rd International Conference on Advanced Computer Theory and Engineering, ICACTE 2010 ; Conference Date: 20 August 2010 Through 22 August 2010.
- [56] D. Witmer, Introduction to computer-mediated communication: A master syllabus for teaching communication technology, *Communication Education* 47 (1998) 162–173. doi:10.1080/03634529809379120.
- [57] R. Greenberg, J. Raphael, J. Keller, S. Tobias, Teaching high school science using image processing: A case study of implementation of computer technology, *Journal of Research in Science Teaching* 35 (1998) 297–327. doi:10.1002/(SICI)1098-2736(199803)35:3<297::AID-TEA4>3.0.CO;2-M.
- [58] J. Larbi-Apau, J. Moseley, Computer attitude of teaching faculty: Implications for technology-based performance in higher education, *Journal of Information Technology Education: Research* 11 (2012) 221–233. doi:10.28945/1711.
- [59] K. Drossel, B. Eickelmann, Teachers' participation in professional development concerning the implementation of new technologies in class: a latent class analysis of teachers and the relationship with the use of computers, ict self-efficacy and emphasis on teaching ict skills, *Large-Scale Assessments in Education* 5 (2017). doi:10.1186/s40536-017-0053-7.
- [60] N. Selwyn, Teaching information technology to the 'computer shy': A theoretical perspective on a practical problem, *Journal of Vocational Education and Training* 49 (1997) 395–408. doi:10.1080/13636829700200023.
- [61] F. Wang, Computer distance virtual experiment teaching application based on virtual reality technology, *International Journal of Emerging Technologies in Learning* 13 (2018) 83–94. doi:10.3991/ijet.v13i04.8472.
- [62] A.-Y. Chen, C.-K. Looi, Teaching, learning and inquiry strategies using computer technology, *Journal of Computer Assisted Learning* 15 (1999) 162–172. doi:10.1046/j.1365-2729.1999.152088.x.

- [63] M. Apple, Teaching and technology the hidden effects of computers on teachers and students, *Educational Policy* 1 (1987) 135–157. doi:10.1177/0895904887001001012.
- [64] Y. Qian, J. Lehman, Using technology to support teaching computer science: A study with middle school students, *Eurasia Journal of Mathematics, Science and Technology Education* 14 (2018). doi:10.29333/ejmste/94227.
- [65] L. Dawes, N. Selwyn, Teaching with the dream machines: The representation of teachers and computers in information technology advertising, *Journal of Information Technology for Teacher Education* 8 (1999) 289–304. doi:10.1080/14759399900200069.
- [66] V. V. Osadchy, I. V. Krasheninnik, Formation of short-cycle curricula content for future software engineers training on the basis of the labour market analysis, *Information Technologies and Learning Tools* 58 (2017) 11–25. URL: <https://journal.iitta.gov.ua/index.php/itlt/article/view/1637>. doi:10.33407/itlt.v58i2.1637.
- [67] V. V. Osadchy, K. P. Osadcha, Modern realities and trends of information and communication technologies development in education, *Information Technologies and Learning Tools* 48 (2015) 47–57. URL: <https://journal.iitta.gov.ua/index.php/itlt/article/view/1252>. doi:10.33407/itlt.v48i4.1252.
- [68] I. G. Bartasevich, A system of criteria for the quality of students' education in higher education, *Bulletin of the Astrakhan state technical university* (2008) 217–219. URL: <https://cyberleninka.ru/article/n/sistema-kriteriev-kachestva-obucheniya-studentov-v-vysshem-uchebnom-zavedenii>.
- [69] V. V. Osadchy, S. V. Symonenko, Foreign language as a means of communicative competence development of future software engineers, *Information Technologies and Learning Tools* 58 (2017) 38–48. URL: <https://journal.iitta.gov.ua/index.php/itlt/article/view/1645>. doi:10.33407/itlt.v58i2.1645.
- [70] P. I. Fedoruk, Adaptive tests: statistical methods of processing results of test knowledge control, *Mathematical Machines and Systems* (2007) 122–138. URL: [http://www.immsp.kiev.ua/publications/articles/2007/2007\\_3,4/Fedoruk\\_034\\_2007.pdf](http://www.immsp.kiev.ua/publications/articles/2007/2007_3,4/Fedoruk_034_2007.pdf).

# Cross platform development vs native development

Nikita A. Shevtsiv, Andrii M. Striuk

Kryvyi Rih National University, 11 Vitalii Matusevych Str., Kryvyi Rih, 50027, Ukraine

## Abstract

The paper analyzes the advantages and disadvantages of cross-platform and native mobile application development. The conditions are highlighted in which native and cross-platform development reveal their advantages. These conditions include the project size, work comfort, popularity, relevance. It was concluded that a beginner developer should start learning from native development, and then try cross-platform.

## Keywords

native development, cross-platform development, React Native, Java, Kotlin, Flutter

## 1. Introduction

In 2010, in the world of mobile development, the main questions were: which to choose Android or iOS? What to learn: Java or Objective-C? Over the course of 10 years, these topics have gradually lost their relevance in the circles of the IT community, because alternatives to these languages have appeared. As Kotlin for Android and Swift for iOS. Parallel to this, a new branch was developing – cross-platform technologies.

In 2020, you can often hear controversy and speculation about which side will take the main part of the mobile app development market: native or cross-platform. Native development involves technologies that are focused on a specific platform. For example, the Java programming language is aimed at developing applications for Android, the brainchild of Apple – Swift for iOS.

For cross-platform development, technologies aimed at both platforms are used. Among them, two of the most promising can be identified: React Native (based on js) from Facebook and Flutter (based on Dart) from Google.

You can ask a mobile developer who has at least a year of experience working with cross-platform technologies and native development (it doesn't matter if it is an Android or iOS developer), he would answer: neither side will ever supplant the other.

But under certain conditions, these parties have clear advantages over each other.

---

CS&SE@SW 2020: 3rd Workshop for Young Scientists in Computer Science & Software Engineering, November 27, 2020, Kryvyi Rih, Ukraine

✉ [nikita0503ua@gmail.com](mailto:nikita0503ua@gmail.com) (N.A. Shevtsiv); [andrey.n.stryuk@gmail.com](mailto:andrey.n.stryuk@gmail.com) (A.M. Striuk)

🌐 <http://mpz.knu.edu.ua/pro-kafedru/vikladachi/224-andrii-striuk> (A.M. Striuk)

🆔 0000-0001-9240-1976 (A.M. Striuk)

© 2020 Copyright for this paper by its authors.  
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).



CEUR Workshop Proceedings (CEUR-WS.org)

## 2. Comparative analysis of cross-platform and native development

### 2.1. Project sizes

Quite often, IT companies with up to 10 employees choose cross-platform. Typically, development teams take on small and medium-sized projects, where development time can take on average from 3 months to a year. As a rule, such projects do not require high technological needs from programmers. Therefore, in these cases cross-platform technologies can be used.

It will be beneficial for both the customer and the company.

The number of application developers can be reduced; most often it will be only one person. On the part of the customer, it will be more comfortable to communicate with one person, rather than explaining to two developers at once how this or that screen should work. In fact, not all people think the same way. And they can understand both the design part and the technical part in different ways, for example, the rules for loading and displaying elements on the screen. Accordingly, the appearance will be not similar on different platforms, and the speed of work will also differ. It will take more time to come to a similar result.

For a small company, being cross-platform has some advantages. Since fewer developers were required, it means that the costs for the application will be less.

It makes no sense to consider the option when one developer first makes an application for one platform, and then for a second one. Since it is obvious that this will take a lot of time and money, so it is not very profitable for both parties.

Will this affect the quality? Applications that do not require a lot of resources from smartphones, for example, simple social network or an online shop, is unlikely to require intervention in low-level processes (Although, in React Native, it is possible to write native modules). So the answer is no. This will not affect the quality and performance of simple applications in any way.

For clients who start their own business and already have a working website, as a project development, cross-platform application development will be cheaper and faster than native.

But this applies to small projects that will not need support, and even more so in further development. If we talk about projects with long development times, constant improvement and expansion of functionality, cross-platform is seriously inferior to native development.

In large projects, a lot of attention is paid to design, and quite often to animation of objects. Despite the fact that, according to the development ideology, this item is noted as a plus, practice shows that the implementation of a non-trivial design for two operating systems gives many bugs. This increases the time to eliminate errors. The UI is displayed differently and the time to adapt is also increased. The most noticeable problems occur with animations, clicks and scrolling – the application may freeze. The user interface is developed in HTML, but you will have to spend months to achieve native platform performance. The native development languages for Android and iOS would be better tools in this case than the cross-platform technology APIs.

You need to develop an interface that is intuitive for both iOS and Android users. Otherwise, an application built according to the IOS Human Interface OS Guidelines will be inconvenient for Android users. Ultimately, you will spend more time improving the user experience.

Long-term projects are not designed to save resources. In this case, customers are looking for performance, quality and reliability. Native development in large-scale projects will be the best solution. For a high-quality application, the project will need developers with serious experience and deep knowledge in the areas of development for Android (Java or Kotlin) and iOS (Objective-C and Swift). Compared to cross-platform technologies, there are 2 times more developers, plus the work of good specialists is more expensive, in other words, the project costs increase. But, it is worth recalling again, large projects involve serious costs, as a result of which, as a rule, this is justified.

Application speed – the compiled project code is optimal for the native platform. The application gets full hardware support for the device and uses multithreading for complex tasks.

With the release of updates to the programming language or hardware functionality, new features will be available immediately. That is, there will be no need to wait for the implementation and adaptation of these updates in cross-platform development technology.

And the most important advantage of native development is the absence of restrictions, and this, in turn, is the main condition for the development of huge projects. An indisputable advantage is that it uses all the functions of the system software as flexibly as possible, such as a camera, microphone, geolocator, accelerometer, calendar, media, at the same time, the application sparingly consumes memory and device battery.

## **2.2. Work comfort**

In addition to the fact that the technology must meet the requirements of the project, it is worth considering a number of factors that are associated with the developer.

The first thing to look out for is the ease of learning. The more popular the technology, the more courses, lessons and articles related to this tool can be found on the Internet. Native development has an advantage in this regard. For example, it is worth considering the Kotlin programming language, it is the official language for developing applications for the Android platform.

Why Google loved Kotlin? On the official website, Android developers write that they have watched the rise of Kotlin in recent years. Google is not shy about describing the language as impressive and concise, more powerful and enjoyable to work with. It has increased performance: the program code on it is, on average, 40% shorter than in other languages, and Kotlin also allows you to avoid some errors in the code. One of the determining factors in Kotlin's popularity with Google is that it is compatible with Java, which is already used in developing applications for Android.

In the future, more libraries and tools will be developed for the language, more teaching materials, it will be easier to find solutions to possible problems.

In other words, the company actively promotes and popularizes the language. Considering that Google is the giant of the IT industry, many companies follow the trend and release a number of courses and other training materials on Kotlin on their own. More and more articles related to the language appear on Habr, on Stack Overflow they solve a huge number of problems using this technology. Thus, the popularity of Kotlin is growing rapidly. It can be concluded that for beginners the entry threshold is greatly reduced and it becomes easier to learn, since there are a lot of materials.

On the other hand, there is the brainchild of the same corporation – Flutter. An incredibly promising technology that Google is promoting in the world of cross-platform development. Flutter, takes a different approach than React Native and Xamarin. It does not turn the source code into native code, which is executed by the platform, but in fact draws a window on the smartphone screen and renders all the elements itself. The language used is Dart, which Google created as an improved version of JavaScript. But as of 2020, the technology is still a bit dry. Due to the fact that Flutter is quite young, the amount of training materials is not as much as we would like, in addition, there are still few additional libraries.

The second point is the threshold of knowledge of entering technology. In other words: the required amount of effort and time that needs to be spent on learning technology so that the knowledge gained is enough to implement a simple project.

From this side, cross-platform comes out a little ahead. Take a Facebook product as an example.

React Native is a JavaScript framework for building natively rendered Android and iOS apps. It is based on Facebook’s React JS library for creating user interfaces. If a programmer has experience in web development, especially in ReactJS, then he can become a developer of applications for mobile devices in no time. Since ReactJS and React Native have 80% similarity in the code base, you can write a business card application for phones almost immediately.

In addition, there is a platform that makes it much easier for beginners to work with React Native.

Expo is a set of tools with which you can write a React Native app in minutes. It includes out-of-the-box tools like Android Studio and XCode configurations, Apple and Google certificate management, push notifications. The Expo development team has developed a wonderful tool called Snack that allows you to write and test code right in the browser, without any pre-installation or setup of development environments. Accordingly, the “threshold of entry” into React Native becomes very low. And thanks to good documentation from Facebook, going from web development to mobile development becomes a breeze. Especially, it is worth noting that this documentation provides materials for those who have experience with native development and recommends working with the React Native CLI. And for those who are new to mobile development, it suggests using the aforementioned Expo CLI [1].

It is worth considering the issue of the popularity and relevance of both sides in more detail.

### 2.3. Popularity

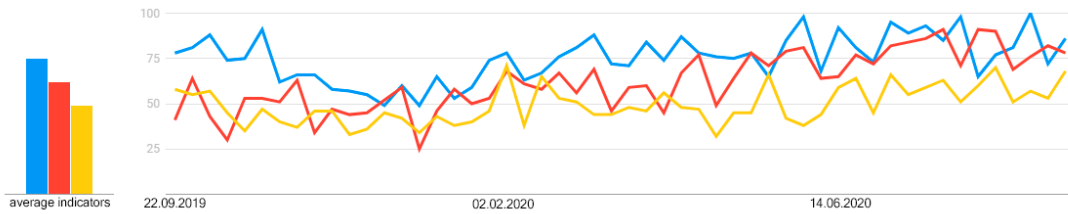
Google Trends [2] service will be used to research popularity.

Consider the popularity of cross-platform development technologies – React Native and Flutter – with the programming language for the Android platform – Kotlin.

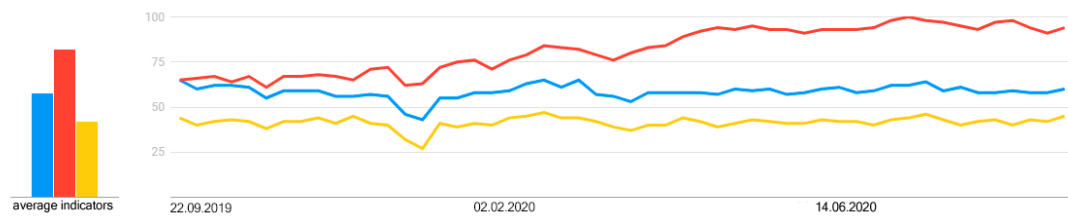
React Native is blue, Flutter is red, Kotlin is yellow (figure 1).

Judging by the graph, React Native has been leading in popularity in Ukraine over the past two years, but Flutter is gaining popularity and even bypasses RN in search statistics at some intervals.

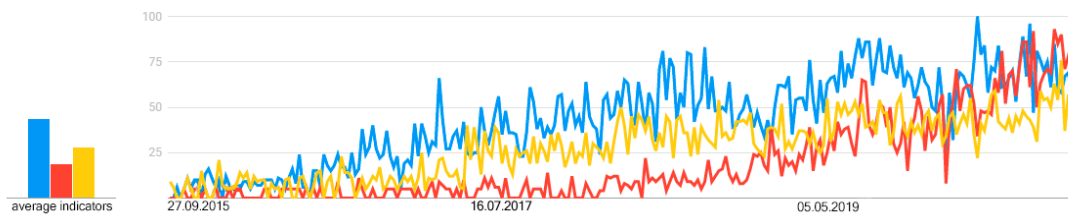
Figure 2 clearly shows how Flutter is gaining popularity every month, more and more. React Native and Kotlin have consistently maintained their popularity in 2019 and 2020.



**Figure 1:** Search statistics for technology names in Ukraine for 2019-2020



**Figure 2:** Search statistics for technology names worldwide for 2019-2020



**Figure 3:** Search statistics for technology names in Ukraine, over the last 5 years

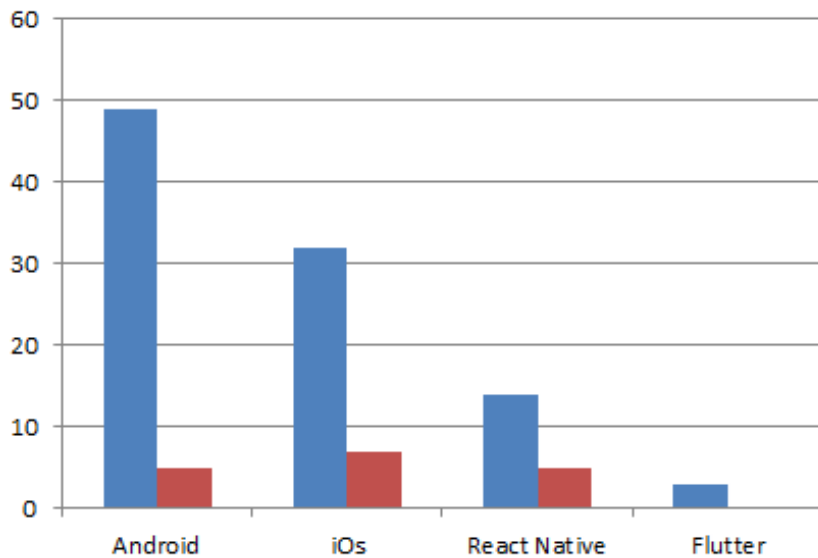
Figure 3 shows the popularity of Google searches for the current list of technologies over the past 5 years in Ukraine. React Native and Kotlin were significantly more popular than Flutter until 2019. But starting in June 2020, Flutter came out on top in the search volume ranking.

The fourth figure shows that over the course of 5 years, React Native has been more popular than the rest of the technologies on the list (despite 2020).

But making a choice, judging only by the number of searches for a programming language on Google, would be the wrong decision. First, the list does not include the Swift programming language for iOS. The fact is that Google Trends under the string “Swift” means not only technology for development, but also as an international interbank system for transferring information and making payments. Secondly, the very popular Java language was not included in the comparison, because it is used not only for developing Android applications, but also for other platforms (web, desktop, etc.). Accordingly, in this comparison, he would have left



**Figure 4:** Search statistics for technology names worldwide, over the past 5 years



**Figure 5:** Job statistics at work.ua

no chance for his opponents. In other words, relying on Google Trends is not a good idea.

But after analyzing many reviews and articles [3, 4] among developers, native development is very popular. Basically, cross-platform development is treated as a tool that only experienced developers use on small projects.

## 2.4. Relevance

The second factor, which is one of the most important in choosing a technology, is its relevance.

Consider the number of vacancies on the following sites: work.ua [5], rabota.ua [6], DOU [7], grc.ua [8].

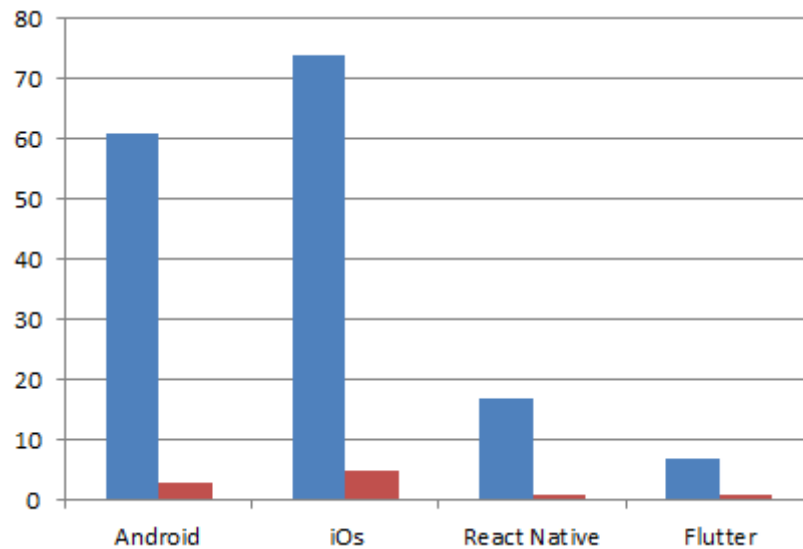
On the graphs, the number of vacancies that are provided for working in the office is marked in blue, and on a remote basis, in red.

Statistics from these sites show that native development in offices is in demand more than cross-platform. But on a remote basis, all technologies have roughly the same number of va-



**Table 1**  
Statistics of vacancies at work.ua

OS / technology	office	remote
Android	49	5
iOS	32	7
React Native	14	5
Flutter	3	0



**Figure 6:** Job statistics on rabota.ua

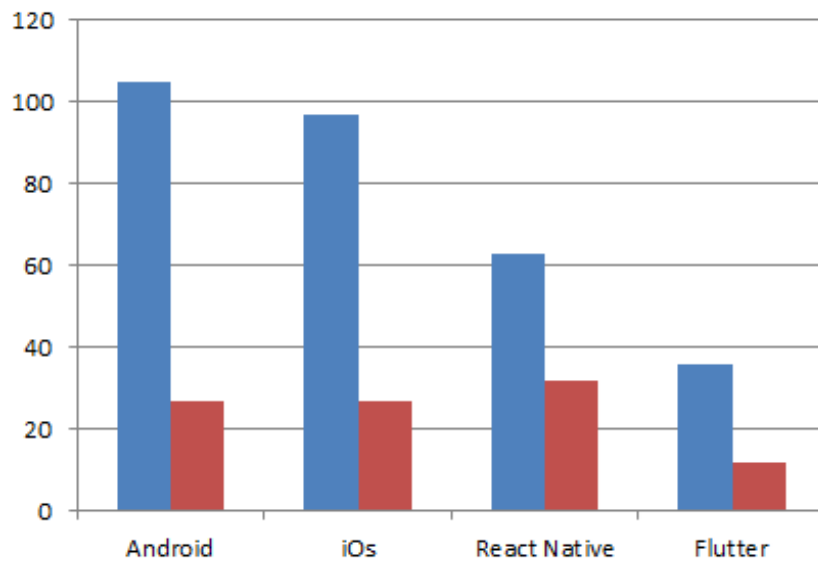
**Table 2**  
Statistics of vacancies at rabota.ua

OS / technology	office	remote
Android	61	3
iOS	74	5
React Native	17	1
Flutter	7	1

cancies.

### 3. Conclusion

In conclusion, it is worth concluding: there is an application of any technology in the labor market, since there are a huge number of ideas from startups to huge and complex projects. It is impossible to give a clear answer which is better: native or cross-platform development.



**Figure 7:** Statistics of vacancies on DOU

**Table 3**

Statistics of vacancies at DOU

OS / technology	office	remote
Android	105	27
iOS	97	27
React Native	63	32
Flutter	36	12

**Table 4**

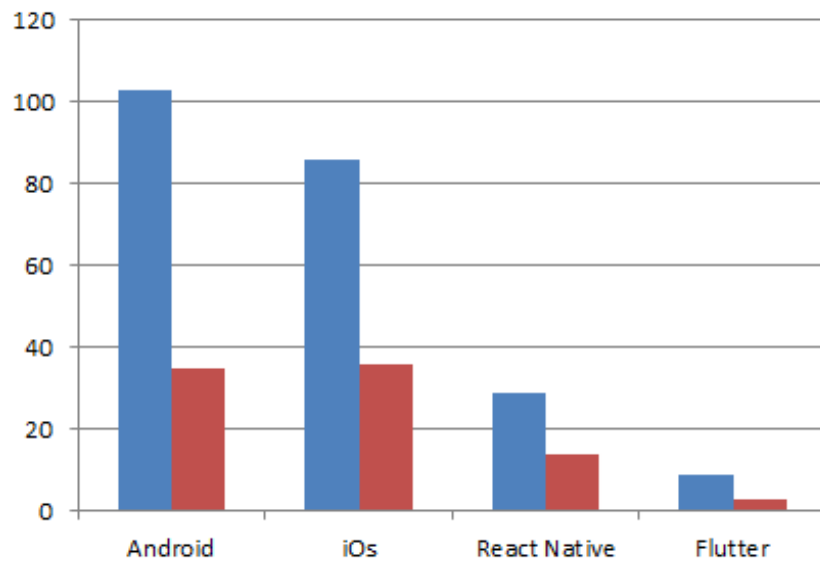
Statistics of vacancies at grc.ua

OS / technology	office	remote
Android	103	35
iOS	86	36
React Native	29	14
Flutter	9	3

Each project has its own technology requirements and is unique. Also, each developer has their own skills and preferences. But, if you are a beginner developer, then you should start learning from native development, then try cross-platform.

## References

[1] Introduction to Expo, 2020. URL: <https://docs.expo.io>.



**Figure 8:** Job statistics on grc.ua

- [2] Google Trends, 2020. URL: <https://trends.google.com/trends/?geo=US>.
- [3] A. Polomodov, Mobile development: native, cross-platform, hybrid, web ..., 2020. URL: <https://apolomodov.medium.com/%D0%BC%D0%BE%D0%B1%D0%B8%D0%BB%D1%8C%D0%BD%D0%B0%D1%8F-%D1%80%D0%B0%D0%B7%D1%80%D0%B0%D0%B1%D0%BE%D1%82%D0%BA%D0%B0-native-cross-platform-hybrid-web-178c355c174>.
- [4] N. Priyatselyuk, Native or cross-platform - what should a novice mobile developer choose? experts answer, 2019. URL: <https://tproger.ru/experts/native-or-crossplatform/>.
- [5] Work.ua, Work.ua is Ukraine's #1 job search site, 2021. URL: <https://www.work.ua/en/?setlp=en>.
- [6] robota.ua, Work in Ukraine. Employment and job search on robota.ua, 2020. URL: <https://rabota.ua>.
- [7] DOU.ua, Community of programmers, 2021. URL: <https://dou.ua>.
- [8] HeadHunter Group of Companies, Work in Ukraine, personnel search and job posting - grc.ua, 2021. URL: <https://grc.ua>.

# WebAR development tools: An overview

Dmytro S. Shepiliev<sup>a</sup>, Yevhenii O. Modlo<sup>d</sup>, Yuliia V. Yechkalo<sup>b</sup>, Viktoriia V. Tkachuk<sup>b</sup>, Mykhailo M. Mintii<sup>a</sup>, Iryna S. Mintii<sup>a,c</sup>, Oksana M. Markova<sup>b</sup>, Tetiana V. Selivanova<sup>a</sup>, Olena M. Drashko<sup>a</sup>, Olga O. Kalinichenko<sup>a</sup>, Tetiana A. Vakaliuk<sup>b,e</sup>, Viacheslav V. Osadchyi<sup>f</sup> and Serhiy O. Semerikov<sup>a,b,c</sup>

<sup>a</sup>Kryvyi Rih State Pedagogical University, 54 Gagarin Ave., Kryvyi Rih, 50086, Ukraine

<sup>b</sup>Kryvyi Rih National University, 11 Vitalii Matusevych Str., Kryvyi Rih, 50027, Ukraine

<sup>c</sup>Institute of Information Technologies and Learning Tools of the NAES of Ukraine, 9 M. Berlynskoho Str., Kyiv, 04060, Ukraine

<sup>d</sup>State University of Economics and Technology, 5 Stephana Tilhy Str., Kryvyi Rih, 50006, Ukraine

<sup>e</sup>Zhytomyr Polytechnic State University, 103 Chudnivska Str., Zhytomyr, 10005, Ukraine

<sup>f</sup>Bogdan Khmelnytsky Melitopol State Pedagogical University, 20 Hetmanska Str., Melitopol, 72300, Ukraine

## Abstract

Web augmented reality (WebAR) development tools aimed at improving the visual aspects of learning are far from being visual and available themselves. This causing problems of selecting and testing WebAR development tools for CS undergraduates mastering in web-design basics. The research is aimed at conducting comparative analysis of WebAR tools to select those appropriated for beginners.

## Keywords

augmented reality software tools, WebAR, A-Frame, AR.js, Three.js, JSARToolKit

---

CS&SE@SW 2020: 3rd Workshop for Young Scientists in Computer Science & Software Engineering, November 27, 2020, Kryvyi Rih, Ukraine

✉ sepilevdmirij@gmail.com (D.S. Shepiliev); eugenemodlo@gmail.com (Y.O. Modlo); uliaechk@gmail.com (Y.V. Yechkalo); viktoriya.tkachuk@gmail.com (V.V. Tkachuk); mykhailo.mintii@gmail.com (M.M. Mintii); irina.mintiy@kdpu.edu.ua (I.S. Mintii); markova@mathinfo.ccjournals.eu (O.M. Markova); vitro090@gmail.com (T.V. Selivanova); lenamd888@gmail.com (O.M. Drashko); olgakalinichenko6@gmail.com (O.O. Kalinichenko); tetianavakaliuk@gmail.com (T.A. Vakaliuk); poliform55@gmail.com (V.V. Osadchyi); semerikov@gmail.com (S.O. Semerikov)

🌐 <https://kdpu.edu.ua/personal/ismintii.html> (I.S. Mintii); <https://kdpu.edu.ua/personal/vstania.html> (T.V. Selivanova); <https://kdpu.edu.ua/personal/omdrashko.html> (O.M. Drashko);

<https://kdpu.edu.ua/personal/ookalinichenko.html> (O.O. Kalinichenko);

<https://sites.google.com/view/neota/profile-vakaliuk-t> (T.A. Vakaliuk);

<https://mdpu.org.ua/new/uk/kafedra-nformatiki-kbneretiki/2990.html> (V.V. Osadchyi);

<https://kdpu.edu.ua/semerikov> (S.O. Semerikov)

🆔 0000-0001-6913-8073 (D.S. Shepiliev); 0000-0003-2037-1557 (Y.O. Modlo); 0000-0002-0164-8365 (Y.V. Yechkalo);

0000-0002-5879-5147 (V.V. Tkachuk); 0000-0002-0488-5569 (M.M. Mintii); 0000-0003-3586-4311 (I.S. Mintii);

0000-0002-5236-6640 (O.M. Markova); 0000-0003-2635-1055 (T.V. Selivanova); 0000-0002-9823-5101 (O.M.

Drashko); 0000-0002-7057-2675 (O.O. Kalinichenko); 0000-0001-6825-4697 (T.A. Vakaliuk); 0000-0001-5659-4774

(V.V. Osadchyi); 0000-0003-0789-0272 (S.O. Semerikov)

© 2020 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).



CEUR Workshop Proceedings (CEUR-WS.org)

## 1. Introduction

Nowadays, augmented, or mixed, reality is a popular technology which is gaining in popularity among educators as well. The basic application of augmented reality (AR) to education is provision of visualization by visualizing computer models of systems and objects using general (a computer equipped with an integrated or external web-camera) and specific (AR goggles, VR helmets) tools, and adapted devices (mobile phones).

The latter are the most widespread devices among participants of the academic process. As of May 2020, there are 8.3 mln users of mobile games, 26.2% of which are 18–24-year olds [1]. According to 2019 data, Ukrainian mobile users at the age of 16-25 are the most frequent users of social networks (92%), message exchange tools (66%), video applications (64%) and mobile games (50%) [2]. Comparison of the two latter sources enables drawing a conclusion that there are from 16 to 16.6 mobile users in Ukraine who are senior school pupils, students of all educational levels or young people. According to the Government Statistics of Ukraine, at the beginning of the 2019–2020 academic year, there are 1.440 mln students of colleges, secondary technical schools, universities, academies and institutes [3], 4.163 mln of which are 16–25-year-olds [3, p. 26].

Comparison of the given statistics indicates that each student has two mobile numbers on average attached to one of the popular stores of mobile applications. It can be either a single mobile phone with two SIM-cards or two mobile phones. As of December 2020, 79.8% of mobile devices in Ukraine had the operating system Android installed [4], while the leading web browser was Chrome (67.8%) [5]. Considering that the third most popular browser, Opera, is built on the Chrome core [6], Chrome accounts for 80%.

In spite of some specific clients for social networking, message exchange, video watching and mobile games, a browser is a universal tool for performing relevant tasks and developing new software in the web-environment quickly. The main advantage of web-clients is their universality as there are Chrome versions for all mobile platforms. Low operating speed of programmes has been the basic disadvantage until recently. Appearance of WebAssembly [7] in March 2017 enabled acceleration of web-programmes and transition of software designed for other platforms into the web-environment.

Access to high-quality libraries of computer vision, image recognition and monitoring (AR-ToolKit and its web-version JSARToolKit5 [8]) is the key to designing WebAR software. At the same time, joint application of JSARToolKit5 and WebGL reveals that:

- markers applied by JSARToolKit5 are a two-dimensional type of a barcode of low visualization,
- visualization of objects by means of WebGL requires in-depth understanding of principles of 3D graphics and greater amounts of a software code.

Thus, the mentioned WebAR development tools are aimed at enhancing visualization of the teaching process, yet they lack visualization and accessibility themselves, this fact indicating the research problem of selecting and testing software tools of WebAR development designed for beginners mastering the basics of web-design - students of lyceums and junior Informatics students.

The following article section are aimed at solving the following task: performing a comparative analysis of WebAR designing tools in order to select those suitable for beginners.

## **2. Review of WebAR designing tools**

### **2.1. Augmented reality: its notion and hardware for Web**

For the recent 60 years, it has been innovative (trendy, interesting, useful, etc.) to program AR, while its application to the web-environment is conditioned by current development of technology. To start this process, one needs an AR-compatible browser like Firefox or Chrome and a text editor (like mcedit or Sublime). Beginners need basic knowledge of web-design (HTML, CSS and JavaScript), while experience of working with web API and GitHub is also applicable.

Hereafter, by AR, we mean an ability of a device (a mobile device or a web browser) to monitor an image or represent a 3D object over the image. The main aim of AR is to show a computer model online in the real-mode space to establish interaction of a user in the real-mode space and a 3D model in the virtual one.

AR can be both marker and markerless. In case of the marker AR, a device tracks a 2D marker: when it is found, a 3D object is visualized. In the markerless variant, a device searches for a flat surface (a table, a floor, etc.) and place a 3D object there.

Using a camera, AR enables visualization of computer-generated objects in games, marketing and other software as in case of arranging furniture in a living room or trying on new clothes before buying. It is a great opportunity for business as it demonstrates a product before a customer really buys it [9].

There are specific devices designed for AR like helmets and headsets enabling a user to plunge into a modelled environment.

AR supplies a real-life world with 3D models controlled by mobile devices anyplace. Virtual reality (VR) makes a user get involved into a modelled world by means of head mounted devices (HMD).

Interactivity in AR and VR software is provided in a similar way. For example, VR actually uses controllers and occasionally hand-tracking enabling a user to interact with 3D objects inside a scene they are situated in.

Major hazards of using HMD in VR include:

- eye strain,
- dizziness and headaches after using HMD.

Unlike VR, AR does not possess any of these hazards for health. Yet, the ability of users to remain concentrated on what they are doing is disturbing, in particular, while using AR, for safety reasons.

Smartphones with iOS (11 and higher under iPhone and iPad control) and Android (7.0 and higher) are the most widespread type of devices for AR.

For web browsers, like Google Chrome and Mozilla Firefox, AR is applicable if they support WebRTC [10] and WebGL [11].

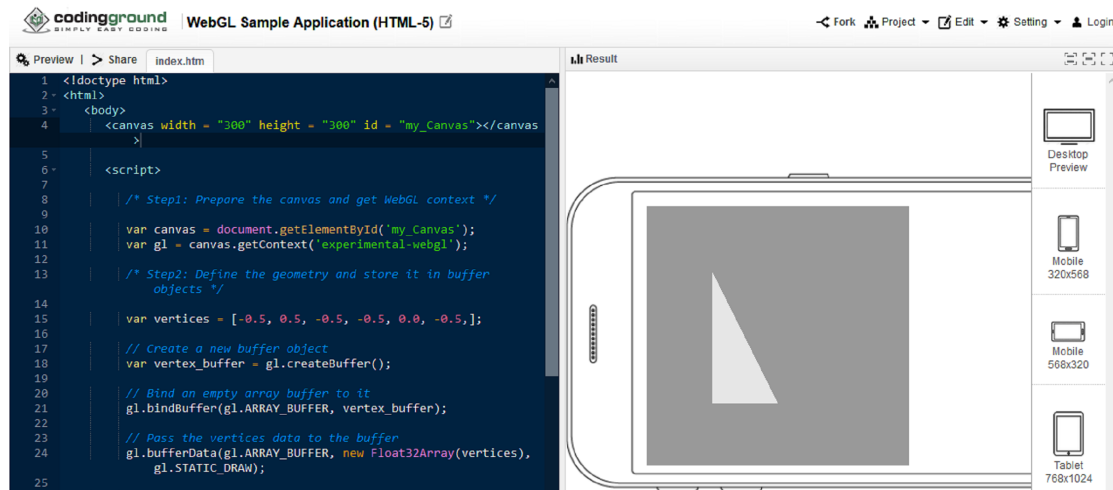


Figure 1: The source code and the result of the simple WebGL programme running.

Microsoft HoloLens is an HMD-like AR headset which is designed currently [12]. Like Google Glass [13], it is designated for corporate use, yet, unlike Glass, it is compatible with Windows 10, not with Android.

## 2.2. Tools of visualizing computer models in the Web

WebGL (OpenGL ES for the Web) is API for 3D graphics in the web browser designed by the Khronos Group Inc [11]. WebGL uses a shader programming language GLSL (OpenGL Shader Language) and is part of an document object model (DOM API) of the browser. All the leading browser designers – Apple (Safari), Google (Chrome), Microsoft (Edge), and Mozilla (Firefox) are members of WebGL Working Group. The current version WebGL 2.0 corresponds to the standard OpenGL ES 3.0 API.

The programme described by WebGL contains both a JavaScript code and a C-like code GLSL. The example of simple software for building a triangle (figure 1) from [14] contains 105 rows of the code comprising five blocks:

1. Preparing a canvas (HTML object canvas) and obtaining a rendering context of WebGL.
2. Defining geometry attributes such as vertexes, indices, etc., and storing them in buffer objects.
3. Designing and compiling software for vertex and fragment shaders.
4. Connecting shader software and buffer objects.
5. Rendering a required image.

Application of WebGL for visualizing computer models in the Web is not the easiest way to do it: the code will account for thousands of rows and require high-level skills in designing 3D graphics by means of OpenGL. Considering the fact that the first research task involves selection of tools for beginners, it is reasonable to use libraries that make application of WebGL simpler.

**Table 1**  
Assessment of efficiency of using libraries for WebGL

Title	Animation	Native audio	Network access	Considering laws of physics	WebGL version	WebXR support	Import	Export	License	General assessment
A-Frame	+	+	-	-	1.0	+	Many formats (3)	HTML, three.js (2)	Free	10
CopperLicht	+	+	-	+	1.0	-	-	-	Free	5
OSG.JS	+	+	-	-	1.0	+	-	-	Free	5
Three.js	+	+	-	-	2.0	+	Many formats (9)	Many formats (4)	Free	19
Verge3D	+	+	-	+	1.0	+	Many formats (3)	glTF (1)	Commercial	9
Clara.io	+	-	-	+	1.0	+	Many formats (5)	Many formats (6)	Commercial	15
Babylon.js	+	+	-	+	2.0	+	Many formats (5)	Many formats (9)	Free	21

Table 1 contains results of assessing efficiency of using JavaScript libraries for WebGL objects.

Considering the fact that the conducted analysis is designated for teaching beginners, there are the best-rated non-commercial libraries determined for working with WebGL – Babylon.js, Three.js and A Frame.

**Babylon.js** [15] enables designing complex 2D-objects by using an essentially smaller code size than that for WebGL: the code size shown in figure 2 is 5-fold smaller than that of the WebGL-code for a simpler scene in figure 1. The designed scene is controlled by simple controlling elements. Unlike other rating participants, Babylon.js is a physical engine widely applied to designing computer games for the Web.

**Three.js** is a free 3D graphics library of general application written in JavaScript. On <https://threejs.org/>, one can find many high-quality demonstrations designed by using it [16]. Its author, Ricardo Miguel Cabello, known also as Mr.doob, is one of the WebGL pioneers. For this reason, this library is often used for designing other libraries.

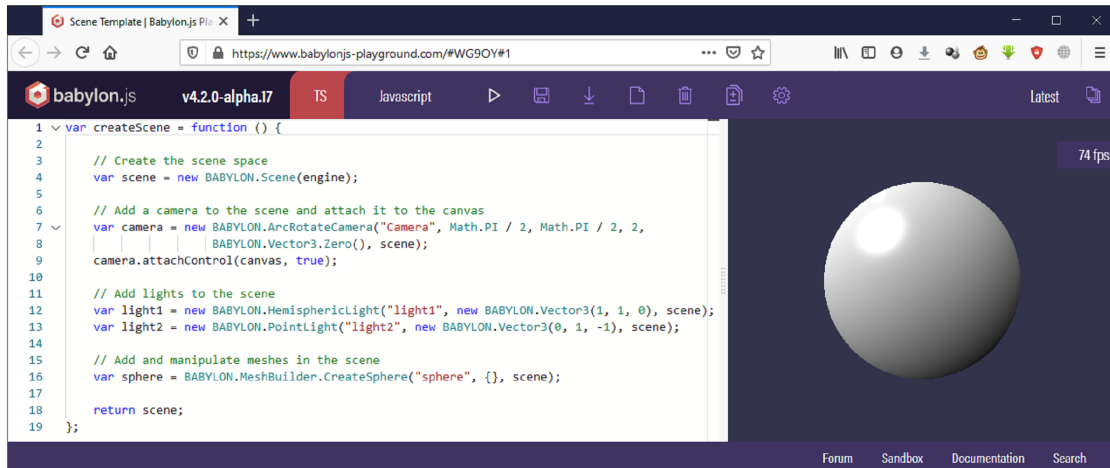
Object designing in Three.js involves three steps:

1. determining object geometry – vectors, colours, etc,
2. determining materials – a rendering method for an object,
3. composing geometry and materials.

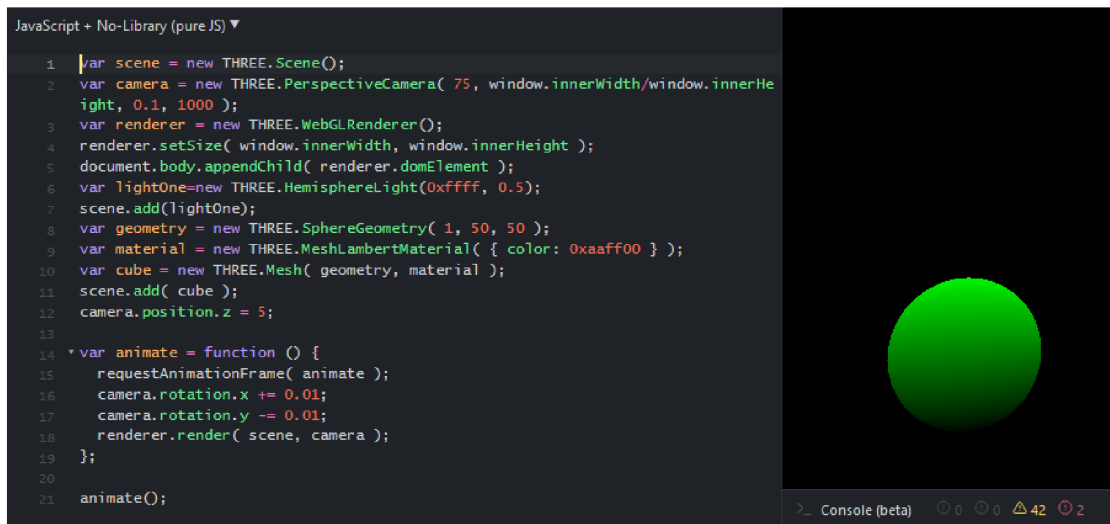
The code size required for designing the 3D scene by Three.js (figure 3) is a bit larger than that of Babylon.js. It is explained by combining in Babylon.js the operations of designing an object and those of adding it to the scene into a single call of a constructor of a relevant class.

**A-Frame** is actually a tool of rapid prototyping: a greater part of its application is a HTML-like code. A-Frame commands are described by tags similar to HTML ones, yet, unlike the





**Figure 2:** The source code and the result of the simple programme running in Babylon.js.



**Figure 3:** The source code and the result of a simple programme running in Three.js.

latter, are interpreted not in a client-side browser, but are a means of accessing JavaScript on the server side [17].

Figure 4 is significant as the A-Frame code size required for designing the same sphere is a three-fold less than that of Babylon.js/Three.js.

Thus, in spite of the third place in the rating of functional opportunities, A-Frame is the leader in terms of visualization and accessibility among the considered libraries for 3D graphics on the Web. Considering the fact that A-Frame is a superstructure over Three.js, it is reasonable to master them in a loop mode starting from A-Frame.



**Figure 4:** The output source code and the result of the simple programme running in A-Frame.

### 2.3. Tools for tracking real-life objects

As [18] contains comparative analysis of the most widespread AR SDK, let us focus on those applicable to designing on the Web – WebAR SDK. Unfortunately, as of January 2021, there are only three free tools, each of which is associated with considered tools of building computer models.

In August 2019, Babylon.js announced **Babylon AR** [19, 20], the project integrating Babylon.js and the computer vision library OpenCV. At present, there is implemented tracking of standard markers similar to QR-codes. In spite of the fact that designers of Babylon.js consider Babylon AR project being on the early stage, web-programmes designed are efficient only on WebXR-supporting devices [21]. This narrows the scope of applying Babylon AR to 8.1 and higher Android devices. In January 2021, there were 74% of such devices among Android ones in Ukraine [22].

Considering the trend of increasing this share, Babylon AR can be considered a promising project of WebAR software development starting from 2021–2022 [22].

ARToolKit, which is 20 years older than Babylon AR, is one of the most applicable AR libraries. **JSARToolKit** (ARToolKit.js) is currently supporting three types of square markers (with arbitrary drawing, a 2D code and markers) and NFT-markers (natural feature tracking) [8]. JSARToolKit is not specific for a WebGL library, yet, it is most frequently used together with Three.js: the example [23] of their joint work developed by Lee Stemkoski is given in figure 5.

The code size required for implementing functionality of the example makes about 100 rows and is much smaller than in case of JSARToolKit. This is due to applying THREE.js – extension of Three.js for designing computer games, this fact making its functionality similar to Babylon.js [24].

Its author, Jerome Etienne [25], is the main designer of the first and the second version of AR.js, which, as of January 2021, has versions adapted to both A-Frame and Three.js. In the same way as A-Frame is a superstructure over Three.js, AR.js is based on JSARToolKit. Thus, this library supports all types of square (marker tracking) and NFT markers (image tracking) as in JSARToolKit, and also enables locating computer models according to their geospatial coordinates (location-based AR) [26].

When applied with A-Frame, the code required for the example similar to that in figure 5



**Figure 5:** The cube on the Hiro marker (joint application of Three.js and JSARToolKit).

takes several rows:

```
<a-scene embedded arjs>
  <a-marker preset="hiro">
    <a-box></a-box>
  </a-marker>
  <a-entity camera></a-entity>
</a-scene>
```

### 3. Conclusions

1. The review of WebAR software design enables recommendations for beginners to master the following combinations of visualization tools of computer models and tools for tracking real-life objects:
  - A-Frame and AR.js is API for fast prototyping most software of which is a HTML-like code. A Frame is applied to designing scenes, objects, animation and other 3D elements in the web browser. AR.js is aimed at tracking markers and enables a scene designed by A-Frame to be rendered directly on the marker;
  - Three.js and ARToolKit.js are used for a deeper level to design software using JavaScript. Three.js uses WebGL to create high-quality 3D scenes in the web browser. JSARToolKit is used for marker tracking and low-level access to data from a camera of a device.

2. Technological conditions of teaching software development with WebAR involve an arbitrary operating system, Internet access, using of cloud storages to deposit materials and availability of at least two classes of devices to test programmes. To master the basics of WebAR software development, it is reasonable to combine A-Frame and AR.js, the scene testing in VR proceeding its AR design. With the scene complexity increasing, JavaScript application to designing A-Frame objects and processing events is accentuated. Designing multi-marker scenes requires different types of matrix markers provided by JSAR-ToolKit, while joint functioning of models attached to different markers requires application of Three.js library. This reveals a dialectic character of teaching WebAR software development: new methods are introduced to overcome problems unsolved by available ones.

## References

- [1] Statista, Mobile Games – Ukraine, 2021. URL: <https://www.statista.com/outlook/211/338/mobile-games/ukraine>.
- [2] D. Elagina, Ukraine: popularity of app categories by age 2018, 2020. URL: <https://www.statista.com/statistics/1023304/ukraine-popularity-app-categories/>.
- [3] Derzhavna sluzhba statystyky Ukrainy, Rozpodil postinoho naseleння Ukrainy za statiu ta vikom na 1 sichnia 2019 roku: statystychnyi zbirnyk (Distribution of the permanent population of Ukraine by sex and age as of January 1, 2019: statistical collection), 2019. URL: [http://database.ukrcensus.gov.ua/PXWEB2007/ukr/publ\\_new1/2020/zb\\_chuselnist%202019.pdf](http://database.ukrcensus.gov.ua/PXWEB2007/ukr/publ_new1/2020/zb_chuselnist%202019.pdf).
- [4] StatCounter, Mobile Operating System Market Share Ukraine, 2021. URL: <https://gs.statcounter.com/os-market-share/mobile/ukraine>.
- [5] StatCounter, Mobile & Tablet Browser Market Share Ukraine, 2021. URL: <https://gs.statcounter.com/browser-market-share/mobile-tablet/ukraine/>.
- [6] Opera Software AS, webkit, 2019. URL: <https://dev.opera.com/tags/webkit/>.
- [7] WebAssembly, 2020. URL: <https://webassembly.org/>.
- [8] artoolkitX, artoolkitx/jsartoolkit5: Javascript ARToolKit v5.x, 2020. URL: <https://github.com/artoolkitx/jsartoolkit5>.
- [9] I. S. Mintii, V. N. Soloviev, Augmented reality: Ukrainian present business and future education, volume 2257, CEUR-WS, 2018, pp. 227–231. URL: <http://ceur-ws.org/Vol-2257/paper22.pdf>, 1st International Workshop on Augmented Reality in Education, AREdu 2018, 2 October 2018.
- [10] Mozilla, individual contributors, WebRTC API – Web APIs, 2021. URL: [https://developer.mozilla.org/en-US/docs/Web/API/WebRTC\\_API](https://developer.mozilla.org/en-US/docs/Web/API/WebRTC_API).
- [11] Mozilla, individual contributors, WebGL: 2D and 3D graphics for the web – Web APIs, 2021. URL: [https://developer.mozilla.org/en-US/docs/Web/API/WebGL\\_API](https://developer.mozilla.org/en-US/docs/Web/API/WebGL_API).
- [12] Microsoft, Mixed reality technology for business, 2021. URL: <https://www.microsoft.com/en-us/hololens>.
- [13] Google, Glass, 2020. URL: <https://www.google.com/glass/start/>.

- [14] Tutorialspoint, WebGL - Sample Application, 2020. URL: [https://www.tutorialspoint.com/webgl/webgl\\_sample\\_application.htm](https://www.tutorialspoint.com/webgl/webgl_sample_application.htm).
- [15] Babylon.js, Babylon.js: Powerful, Beautiful, Simple, Open - Web-Based 3D At Its Best, 2021. URL: <https://www.babylonjs.com/>.
- [16] R. Cabello, three.js – JavaScript 3D library, 2021. URL: <https://threejs.org/>.
- [17] Supermedium, A-Frame – Make WebVR, 2020. URL: <https://aframe.io/>.
- [18] O. V. Syrovatskyi, S. O. Semerikov, Y. O. Modlo, Y. V. Yechkalo, S. O. Zelinska, Augmented reality software design for educational purposes, volume 2292, CEUR-WS, 2018, pp. 193–225. URL: <http://ceur-ws.org/Vol-2292/paper20.pdf>, student Workshop on Computer Science and Software Engineering, CS&SE@SW 2018, 30 November 2018.
- [19] Babylon.js, Babylon AR Preview CDN, 2019. URL: <https://ar.babylonjs.com/>.
- [20] J. Murray, Babylon.js, Babylon AR, 2019. URL: <https://medium.com/@babylonjs/babylon-ar-7823ab4a80c1>.
- [21] Mozilla, individual contributors, WebXR Device API – Web APIs, 2021. URL: [https://developer.mozilla.org/en-US/docs/Web/API/WebXR\\_Device\\_API](https://developer.mozilla.org/en-US/docs/Web/API/WebXR_Device_API).
- [22] StatCounter, Mobile Android Version Market Share Ukraine, 2021. URL: <https://gs.statcounter.com/android-version-market-share/mobile/ukraine>.
- [23] L. Stemkoski, Basic Cube, 2018. URL: <https://stemkoski.github.io/AR-Examples/hello-cube.html>.
- [24] J. Etienne, THREE.js Game Extensions for Three.js, 2017. URL: <https://www.threejsgames.com/extensions/>.
- [25] J. Etienne, jeromeetienne/threex: Game Extensions for three.js, 2017. URL: <https://github.com/jeromeetienne/threex>.
- [26] AR.js organization, AR.js-org/AR.js: Image tracking, Location Based AR, Marker tracking. All on the Web, 2020. URL: <https://github.com/AR-js-org/AR.js>.

# Quantum enhanced machine learning: an overview

Pavlo V. Zahorodko<sup>a</sup>, Yevhenii O. Modlo<sup>d</sup>, Olga O. Kalinichenko<sup>a</sup>, Tetiana V. Selivanova<sup>a</sup> and Serhiy O. Semerikov<sup>a,b,c</sup>

<sup>a</sup>Kryvyi Rih State Pedagogical University, 54 Gagarin Ave., Kryvyi Rih, 50086, Ukraine

<sup>b</sup>Kryvyi Rih National University, 11 Vitalii Matusevych Str., Kryvyi Rih, 50027, Ukraine

<sup>c</sup>Institute of Information Technologies and Learning Tools of the NAES of Ukraine, 9 M. Berlynskoho Str., Kyiv, 04060, Ukraine

<sup>d</sup>State University of Economics and Technology, 5 Stephana Tilhy Str., Kryvyi Rih, 50006, Ukraine

## Abstract

Machine learning is now widely used almost everywhere, primarily for forecasting. The main idea of the work is to identify the possibility of achieving a quantum advantage when solving machine learning problems on a quantum computer.

## Keywords

machine learning, quantum computing, quantum software engineering

## 1. Introduction

Traditionally, quantum computing is defined as a type of nonclassical computing that operates on the quantum state of subatomic particles, which represent information as elements denoted as quantum bits (qubits). A qubit can represent all possible values simultaneously (superposition) until read. Qubits can be linked with other qubits, a property known as entanglement. Quantum algorithms manipulate linked qubits in their undetermined (entangled) state, a process that can address problems with vast combinatorial complexity [1], reaching “quantum supremacy”.

Identifying potential applications for quantum computing, Kasey Panetta points out that they “will be narrow and focused, as general-purpose quantum computing will most likely never be economical” [2]. In his opinion, quantum computing could enable breakthroughs by machine learning, finance, healthcare, creation of new materials, artificial intelligence (which requires 100s – 1000s qubits), chemistry and biochemistry (100–200 qubits). In particular, for finance, quantum computing could enable faster, more complex Monte Carlo simulations (for

---

CS&SE@SW 2020: 3rd Workshop for Young Scientists in Computer Science & Software Engineering, November 27, 2020, Kryvyi Rih, Ukraine

✉ pavelzagorodko@outlook.com (P.V. Zahorodko); eugenemodlo@gmail.com (Y.O. Modlo); olgakalinichenko6@gmail.com (O.O. Kalinichenko); vitro090@gmail.com (T.V. Selivanova); semerikov@gmail.com (S.O. Semerikov)

🌐 <https://kdpu.edu.ua/personal/ookalinichenko.html> (O.O. Kalinichenko);

<https://kdpu.edu.ua/personal/vstania.html> (T.V. Selivanova); <https://kdpu.edu.ua/semerikov> (S.O. Semerikov)

🆔 0000-0003-2037-1557 (Y.O. Modlo); 0000-0002-7057-2675 (O.O. Kalinichenko); 0000-0003-2635-1055 (T.V.

Selivanova); 0000-0003-0789-0272 (S.O. Semerikov)

© 2020 Copyright for this paper by its authors.  
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).



CEUR Workshop Proceedings (CEUR-WS.org)

example, trading, trajectory optimization, market instability, price optimization and hedging strategies) and machine learning methods, which in the general case are reduced to problems of finding the extremum of a multidimensional function along the nonlinear response surface.

Currently, computing devices capable of performing quantum computing (quantum computers) are available for consumers of computing services using the QCaaS (quantum computing as a service) model. As of June 2020, the maximum number of qubits available for simultaneous use does not exceed 60, which is significantly less than the number required to achieve “quantum supremacy”. This raises the problem of investigating the possibilities of quantum programming for machine learning tasks implementation, namely, the use of machine learning algorithms, implemented by the quantum programming language, to analyze traditional data and compare the performance of quantum and von-neumanns implementations at the present stage of their development.

## 2. Fundamentals of Quantum Software Engineering

### 2.1. Basic research concepts

*Quantum computer* is a computing device using quantum-mechanical phenomena (superposition, entanglement, etc.) for data transmission and processing.

*Quantum programming* is a software development process for quantum computer.

“Classical” applications of quantum computers (by Richard Feynman) – modeling complex [many-particle physical] systems: Zalka and Wiesner’s algorithm.

“New” applications of quantum computers are tasks that require enumerating a large number of options: Grover’s algorithm (general task), Shor’s algorithm (factorization), Abrams and Lloyd’s algorithm (identification of periodic properties), etc.

*Quantum machine learning* is an application of machine learning algorithms for quantum data analysis.

*Quantum-enhanced machine learning* is the use of machine learning algorithms implemented in the quantum programming language for the analysis of traditional data.

*Software Engineering* is a systematic application of engineering approaches to the design, implementation, testing and documenting of software.

### 2.2. Concept of Quantum Software Engineering

The first systems presentation of the Quantum Software Engineering concept was made by John Clark and Susan Stepney in 2002 [3]. Researchers believe that quantum computing cannot be effectively implemented in the traditional computer Von Neumann architecture, the mathematical model of which is the Turing machine. The authors [4] refer to the main challenges that Quantum Software Engineering will face in 2020:

- the question of what a quantum programming language should be – an extension of traditional languages, a logical programming language in a low-level programming language or a language that implements a new paradigm,
- the need to develop compilers for quantum programming languages,

- the need to develop new quantum algorithms and define the classes of traditional algorithms that can be quantised,
- feasibility of developing quantum computer simulators for use on traditional computer systems,
- despite the fact that quantum execution is in principle unobservable, debugging and testing techniques are necessary for quantum programming languages,
- quantum algorithms require visualization for their understanding, design, and implementation.

The criteria and success indicators of Quantum Software Engineering proposed by John Clark and Susan Stepney are summarized in table 1.

In 2020, Quantum Software Engineering includes such components [5]:

- Paradigms for developing quantum software
- Quantum software design
- Quantum software testing
- Quantum software verification
- Quantum software coding practices
- Quantum software reuse
- Quantum software experimentations
- Quantum software execution
- Industrial applications
- Empirical evaluations

In February 2020, at QANSWER 2020: 1st International Workshop on the QuANtum Software Engineering & pRogramming, the Talavera Manifesto for Quantum Software Engineering and Programming [6] was adopted, containing a set of principles and commitments:

*Quantum Software Engineering*

- *is agnostic regarding quantum programming languages and technologies;*
- *embraces the coexistence of classical and quantum computing, and advocates the use of reengineering techniques to integrate new quantum algorithms with the existing classical information systems. Reverse engineering techniques are also needed to parse and abstract quantum program information that is to be integrated into classical programs;*



- *supports the management of quantum software development projects*, delivering quantum software that fulfils the initial business goal and requirements, while at the same time ensuring that quality, time, and cost constraints are being properly observed; methodologies for developing quantum programs must be created or adapted from the existing ones; effort estimation methods for quantum software development need to be provided as well;
- *considers the evolution of quantum software*: quantum software should be maintained and evolved from inception to removal, and quantum software evolution must be handled throughout the whole quantum software lifecycle;
- *aims at delivering quantum programs with desirable zero defects*: it is in charge of defining and applying testing and debugging techniques to quantum programs in such a way that most defects can be detected and solved before the program is released;
- *assures the quality of quantum software*: quality management for both process and product are essential if quantum software with expected quality levels is to be produced; since we cannot improve what we cannot measure, new metrics for quantum programs and quantum processes have to be developed;
- *promotes quantum software reuse*, helping development teams to share, index, and find quantum software that can be reused: this requires study of design and architectural patterns for quantum programs, facilitate technical communication, and work on creating libraries of reference examples and application demonstrations;
- *addresses security and privacy by design*: quantum information systems must be secure and guarantee the privacy of data and of users from the initial phases of quantum software development, i.e., by design;
- *covers the governance and management of software*: managers should be aware of the particular processes, organizational structures, principles, policies and frameworks, information, culture, ethics and behaviour, people, skills and competences, as well as the services, infrastructure and applications that are associated with quantum software and that are (or should be) provided by organizations.

The authors of the manifesto separately appeal to educators with a request to integrate quantum software engineering in curricula within the existing software engineering degrees and/or courses in this or other disciplines, and clearly specify which competences and skills are required for future quantum software engineers [6].

### **2.3. Quantum Software Engineering tools**

The execution of quantum programs on personal computer equipment is difficult to access due to its lack of prevalence, so for more than a quarter-century, quantum simulators – software tools that simulate quantum circuits – have been the main means of their execution. The first mention of QCaaS (Quantum Computing as a Service) occurs only in 2015 in the article [7] by Mijanur Rahaman and Md. Masudul Islam.

The world's largest QCaaS providers:

- D-Wave Systems Inc. (Canada) – SDK Ocean [8] (Python, C++),
- International Business Machines Corporation (USA) – SDK ProjectQ [9] (Python), Qiskit [10] (Python),
- Cambridge Quantum Computing Limited (Great Britain) – SDK t|ket> [11] (Python),
- QC Ware, Corp. (USA) – SDK Forge (Python),
- StationQ – Microsoft (USA) – SDK LIQUi|> [12] (F#), Microsoft Quantum Development Kit [13] (F#),
- Rigetti Computing (USA) – SDK Forest [14] (Python).

Thus, the main programming language for cloud access to quantum computing is Python. Another criterion for choosing a QCaaS vendor is computing power, measured in qubits. This indicator is the largest in D-Wave Advantage – 5000 (in clusters of 8) qubits based on quantum annealing, which narrows the scope of its application to solving optimization problems, which boil down to finding the ground state for a set of spins. For universal quantum computers on quantum circuits, the number of qubits is significantly lower and today (June 2020) is the highest in IBM Q 53 (53 qubits) and Google Bristlecone (72 qubits). Unfortunately, Google's Quantum Computing Playground [15] is a browser-based quantum simulator, and there is no open cloud access to Google's Bristlecone. For cloud access to IBM Q, you can use both their library – Qiskit, and a third-party – ProjectQ. Considering that the highest level of specialization is provided by its own SDK, Qiskit was chosen for further work.

### 3. Quantum-enhanced machine learning

#### 3.1. Quantum models of machine learning

Srinivasan Arunachalam and Ronald de Wolf in [16] offer three main quantum learning models:

1. *Quantum exact learning* based on membership queries to find the most accurate unknown function (quantum approximation problem). The efficiency of quantum algorithms in relation to classical ones in this case depends on how the learning efficiency is measured. If the measure of efficiency is the training time, then there are such classes of functions for which quantum algorithms are much faster than classical ones, assuming that the queries implementation in a quantum superposition is possible.
2. *Quantum Probably Approximately Correct (PAC) learning* to find an unknown function over a set of samples (quantum supervised learning). The difference between quantum PAC learning and classical learning is that the dataset can be in a state of quantum superposition.
3. *Quantum agnostic learning* to search for the  $(n + 1)$ -th bit, which is a continuation of a sequence with  $n$  bits (quantum prediction task).

The authors point to three types of complexity that arise when applying quantum learning models [16]:

1. query complexity of quantum exact learning: the number of quantum membership queries needed to exactly learn a target concept can be polynomially smaller than the number of classical membership queries, but not much smaller than that,
2. sample complexity: for the distribution-independent models of PAC and agnostic learning, quantum examples give no significant advantage over classical random examples: for every concept class, the classical and quantum sample complexities are the same up to constant factors. In contrast, for some fixed distributions (e.g., uniform) quantum examples can be much better than classical examples,
3. time complexity: there exist concept classes that can be learned superpolynomially faster by quantum computers than by classical computers, for instance based on Shor's or Simon's algorithm.

In the case of applying quantum machine learning models to the analysis of traditional data, we are talking about quantum-enhanced machine learning. Frank Phillipson [17] defines three main benefits of quantum machine learning:

- improving runtime (for example with a quantum hybrid Helmholtz machine)
- learning capacity improvements (for example with a quantum Hopfield neural network)
- learning efficiency improvements: less training information or simpler models needed to produce the same results or more complex relations can be learned from the same data

Various methods can be applied to increase the efficiency of training, one of which is variational quantum circuits – VQC [17].

Evidence of the intensity of quantum-enhanced machine learning development is the fact that the systematic review of the problem in 2016, carried out by Peter Wittek in [18], today (November 2020) is already considered as a classic, and that is indicated by the co-author in a new review [19].

Vedran Dunjko and Peter Wittek also highlight such perspective directions in the development of quantum machine learning in general:

- supervised and unsupervised learning: continuous-variable quantum neural networks, quantum convolutional neural networks, quantum algorithms for feedforward neural networks, Bayesian deep learning, sublinear quantum algorithms for training linear and kernel-based classifiers,
- reinforcement learning: quantum algorithms for solving dynamic programming problems (including hidden quantum Markov models), quantum gradient estimation.

The authors conclude that “the entire field of “genuinely quantum” machine learning (where the data itself is quantum) is still finding its right place and full recognition. Perhaps as quantum technologies mature, and problems of quantum learning become genuinely practical, the field

will crystallize and grow. ... In summary, QML [quantum machine learning] is diverse, growing, inclusive, and it is rich in open questions. ... Capturing all the QML trends, which will in the end be central is, for the time being, an impossible task – and, in a way, this is the key message of this note” [19].

### 3.2. An overview of quantum-enhanced machine learning tools in Qiskit

Qiskit provides the ability to develop quantum software both at the quantum circuits level using OpenQASM [20] and at a high level of abstraction using Python in a Jupyter notebook. The main components of the library are:

- quantum circuits modeling tools (Terra),
- implementation of standard quantum algorithms (Aqua – Algorithms for QUantum Applications), in particular, for solving optimization tasks
- cloud quantum computing tools (Aer),
- tools for simulating quantum noise (Ignis).

Aqua includes modules for research in finance (`qiskit.finance`), machine learning (`qiskit.ml`), optimization (`qiskit.optimization`) and chemistry (`qiskit.chemistry`) [20].

The machine learning module contains standard datasets and ways to access custom. Various optimization algorithms can be used to process them:

- `ADMMOptimizer` – an implementation of the ADMM-based heuristic (ADMM – alternating direction method of multipliers)
- `CobylaOptimizer` – the SciPy COBYLA optimizer (COBYLA – Constrained Optimization BY Linear Approximation)
- `CplexOptimizer` – the CPLEX optimizer for linear, integer and quadratic programming tasks
- `GroverOptimizer` – uses Grover Adaptive Search (GAS) to find the minimum of a QUBO function (QUBO – quadratic unconstrained binary optimization)
- `MinimumEigenOptimizer` – minimum eigen solvers
- `RecursiveMinimumEigenOptimizer` – a meta-algorithm that applies a recursive optimization

The `qiskit.aqua.components.optimizers` module offers a set of algorithms for local (Analytic quantum gradient descent optimizer, constrained optimization by linear approximation optimizer, Nelder-Mead optimizer, Nakanishi-Fujii-Todo algorithm, Powell optimizer, truncated Newton optimizer, etc.) and global optimizations (controlled random search with local mutation optimizer, evolutionary optimizer, etc.). It is advisable to use quantum support vector machine (QSVM) and variational quantum classifier (VQC) algorithms to solve classification tasks.

## 4. Conclusions

1. The core of Quantum Software Engineering is quantum programming – the process of developing programs for a quantum computer: a computing device that uses the phenomena of quantum mechanics to process data. Due to the low level of availability of such devices, it is advisable to access them under QCaaS model (quantum computing as a service). The conducted review of Quantum Software Engineering tools provided an opportunity to single out their main classes (quantum simulators, libraries, visualizers and cloud quantum services) and recommend using IBM Q as a hardware platform for quantum computing, Qiskit as a library of quantum algorithms, Python as a programming language and IBM Quantum Experience as QCaaS Provider.
2. The use of machine learning algorithms for the analysis of quantum data can be described by three quantum machine learning models (quantum exact learning, quantum Probably Approximately Correct learning and quantum agnostic learning), in the application of which there are three types of difficulties associated with the query complexity of quantum exact learning, quantum the intricacy of datasets and the sensitivity of quantum algorithms to them. A prospective direction in the machine learning development is the use of quantum learning models for analyzing traditional data, the implementation of which in Qiskit Aqua 0.7.3 is still a limited solution to classification tasks.

## References

- [1] Gartner, Quantum Computing Gartner Glossary, 2021. URL: <https://www.gartner.com/en/information-technology/glossary/quantum-computing>.
- [2] K. Panetta, The CIO's Guide to Quantum Computing, 2019. URL: <https://www.gartner.com/smarterwithgartner/the-cios-guide-to-quantum-computing/>.
- [3] J. Clark, S. Stepney, Quantum software engineering, in: Workshop on Grand Challenges for Computing Research, e-Science Institute, Edinburgh, 2002. URL: <http://web.archive.org/web/20200721161705/http://www.ukcrc.org.uk/press/news/call/a5.cfm>.
- [4] C.-H. Chenf, L.-Y. Wei, New entropy clustering analysis method based on adaptive learning, in: P. P. Wang (Ed.), Information Sciences 2007: Proceedings of the 10th Joint Conference, Salt Lake City, Utah, USA, 18–24 July 2007, 2007, pp. 1196–1202. doi:10.1142/9789812709677\_0169.
- [5] Q-SE2020, First International Workshop on Quantum Software Engineering (Q-SE 2020) co-located with ICSE 2020, 2021. URL: <https://q-se.github.io/qse2020/>.
- [6] M. Piattini, G. Peterssen, R. Perez-Castillo, J. L. Hevia, M. A. Serrano, G. Hernández, I. G. R. de Guzmán, C. A. Paradela, M. Polo, E. Murina, L. Jiménez, J. C. Marqueno, R. Gallego, J. Tura, F. Phillipson, J. M. Murillo, A. Niño, M. Rodríguez, The Talavera Manifesto for Quantum Software Engineering and Programming, CEUR Workshop Proceedings 2561 (2020) 1–5.
- [7] M. Rahaman, M. M. Islam, A Review on Progress and Problems of Quantum Computing as aService (QCaas) in the Perspective of Cloud Computing, Global Journal of Computer Science and Technology: B Cloud and Distributed 15 (2015) 15 – 18. URL: [https://globaljournals.org/GJCST\\_Volume15/3-Cloud-Data-Storage.pdf](https://globaljournals.org/GJCST_Volume15/3-Cloud-Data-Storage.pdf).

- [8] D-Wave Systems Inc, D-Wave Ocean Software Documentation, 2021. URL: <https://ocean.dwavesys.com/>.
- [9] D. Steiger, T. Häner, ProjectQ – Open Source Software for Quantum Computing, 2017. URL: <https://projectq.ch/>.
- [10] Qiskit, Qiskit, 2021. URL: <https://qiskit.org/>.
- [11] Cambridge Quantum Computing, Technology, 2020. URL: <https://cambridgequantum.com/technology/>.
- [12] Microsoft, Language-Integrated Quantum Operations: LIQUI|>, 2016. URL: <https://www.microsoft.com/en-us/research/project/language-integrated-quantum-operations-liqui/>.
- [13] Microsoft, Microsoft Quantum Documentation and Q# API Reference - Microsoft Quantum, 2021. URL: <https://docs.microsoft.com/en-us/quantum/>.
- [14] Rigetti Computing, Rigetti QCS, 2020. URL: <https://qcs.rigetti.com/sdk-downloads>.
- [15] Google, Quantum Computing Playground, 2016. URL: <http://www.quantumplayground.net>.
- [16] S. Arunachalam, R. de Wolf, A Survey of Quantum Learning Theory, 2017. [arXiv:1701.06806](https://arxiv.org/abs/1701.06806).
- [17] F. Phillipson, Quantum Machine Learning: Benefits and Practical Examples, CEUR Workshop Proceedings 2561 (2020) 51–56.
- [18] P. Wittek, Quantum Machine Learning: What Quantum Computing Means to Data Mining, Elsevier Insights, Academic Press, San Diego, 2016.
- [19] V. Dunjko, P. Wittek, A non-review of Quantum Machine Learning: trends and explorations, Quantum Views 4 (2020) 17. doi:10.22331/qv-2020-03-17-32.
- [20] M. Pistoia, J. Gambetta, Qiskit Aqua – A Library of Quantum Algorithms and Applications, 2018. URL: <https://medium.com/qiskit/qiskit-aqua-a-library-of-quantum-algorithms-and-applications-33ecf3b36008>.

**Table 1**

The criteria and success indicators of Quantum Software Engineering (according to [3])

Criteria	Indicators
It arises from scientific curiosity about the foundation, the nature or the limits of a scientific discipline	Quantum computation has broadened the fundamental limits of computer science and software engineering
The ability to create new engineering solutions	The physical infrastructure is constantly evolving, each solution is new
Technological continuity	The existence of high level languages and development techniques that can be used by computer scientists and software engineers with only the same style of training they receive today (so, no need to teach the fundamentals of quantum mechanics to all)
Research community support	Support for all interested in new computing paradigms and new levels of computing power
International character of research	This is a new fundamental area of software engineering
It is generally comprehensible, and captures the imagination of the general public, as well as the esteem of scientists in other disciplines	It is not generally understood, but is known for its worldwide interpretation
The problem has a long-standing statement, but has not yet been resolved	Formulated by Richard Feynman in the late 1970s
It promises to go beyond what is initially possible, and requires development of understanding, techniques and tools unknown at the start of the project	Problems exist on every level, from developing a whole new conceptual paradigm, to building intellectual and simulation tools
It calls for planned co-operation among identified research teams and communities	Research is needed in a number of areas (languages, algorithms, tools, simulation, visualisation, etc.)
It encourages and benefits from competition among individuals and teams, with clear criteria on who is winning, or who has won	There need not be a single “winner”, diversity of solutions should be encouraged, as in classical software engineering, to be applicable to a range of application domains
It decomposes into identified intermediate research goals, whose achievement brings scientific or economic benefit, even if the project as a whole fails	There are several components of the problem that can be explored in parallel
It will lead to radical paradigm shift	Quantum computing is a radical paradigm shift

# Development of an automated system for conducting, checking and evaluating programming competitions

Bohdan V. Hrebeniuk, Olena H. Rybalchenko

*Kryvyi National University, 11 Vitalii Matusevych Str., Kryvyi Rih, 50027, Ukraine*

## Abstract

The paper analyzes the existing platforms for conducting programming contests. Possible approaches are analyzed for creating isolated environments and running participants' solutions, advantages and disadvantages of both approaches are highlighted. Requirements for the user interface are defined that must provide quick and convenient work in the system; the system was planned and developed. It was concluded that designed system has a potential for conducting contests and further development.

## Keywords

competitive programming, isolated environment, virtualization, containerization, REST

## 1. Introduction


Competitive programming has been maintaining a dynamic development rate for several decades. Year by year, more and more teams take part in local, regional and international contests, therefore making the needs of both contest organizers and participants grow.


Kryvyi Rih National University supports the development of competitive programming – local contests are held, and the university annually hosts the qualification phase of the international student contest under the aegis of ACM/ICPC. To maintain the level of students and teachers competence, it is necessary to have our own flexible system, since existing analogues have certain disadvantages.


The aim of the work is to study and analyze the process of creating isolated environments, design and develop a flexible system for conducting, checking and evaluating programming contests.


---

*CS&SE@SW 2020: 3rd Workshop for Young Scientists in Computer Science & Software Engineering, November 27, 2020, Kryvyi Rih, Ukraine*

 bogdan020699@gmail.com (B.V. Hrebeniuk); rybalchenko@knu.edu.ua (O.H. Rybalchenko)

 0000-0002-0423-8476 (B.V. Hrebeniuk); 0000-0001-8691-5401 (O.H. Rybalchenko)

 © 2020 Copyright for this paper by its authors.  
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)



## **2. Development of a system for conducting, checking and evaluating programming competitions**

### **2.1. Analysis of existing systems**

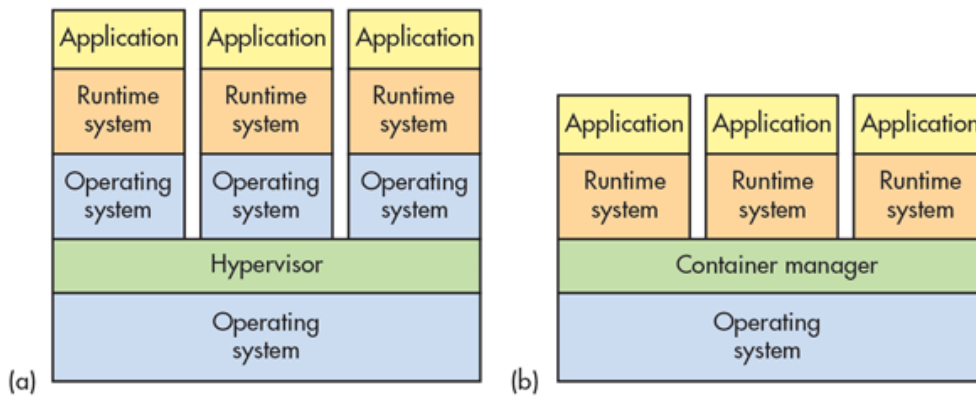
One of the first systems with automatic checking of tasks is UVa Online Judge, an online system of the University of Valladolid with an installed task archive. The archive is constantly being updated, but it is impossible to create tasks yourself using this system – any system updates (uploading new tasks, supporting new programming languages, changing the rules for creating a rating) are possible only with permission and active interaction with developers. Additionally, downloading the software and installing it for use are prohibited due to the product license, because the system uses the “SaaS” (Software as a Service) model [1]. Under this model, the software is not installed on users’ computers, but on the provider’s servers, and the user does not have any access to the system. This disadvantage makes the system unsuitable for long-term contests due to the fact that the terms of the service provider can change at any moment and the user (in our case, the organizer of the contest) will not be able to do anything [2].

One of the most successful systems in history was PCSM2, which was formed by combining the APPES and PCMS systems and is characterized by the flexibility of the internal structure. Software units have low connectivity due to the fact that they rely on high-level abstractions of each other. This makes the PCSM2 system easy to upgrade on a programming level. But this flexibility has made it difficult to customize the server, as most of the business logic is described by configuration files. Some parts of the system are still being developed, so to set them up, you need to add and connect specialized modules that only the authors of the system can perform. Another serious drawback is that the PCSM2 server supports only the Windows operating system, which makes it difficult to run the system on servers that use Linux [3].

The qualification phases of the ACM/ICPC contests are managed by the ejudge system [4], which is significantly different from PCSM2. The system was created using the C programming language, it supports only the Linux operating system, has a web interface for administration and detailed documentation, but also has a number of disadvantages. The most important of these is the inefficient allocation of resources between the system and processes with user’s solutions. As a result, incorrect verdicts may occur due to exceeding the time limit. The system also doesn’t support scaling out – you can’t connect a new machine to an existing one. The only way to improve parallel code execution is to increase the number of processes on the server, which can lead to competition for process access to RAM. An equally important factor is linking the system to the operating system. This dependency makes it impossible to deploy the project on servers that do not support Linux.

### **2.2. Analysis of approaches of running participants’ code**

Automation of contest processes has become a necessary condition for such events, because with the increase in the number of teams and tasks, the required amount of time for receiving solutions, checking them, and summing up results has also significantly increased. And, although such systems have existed for decades, with the introduction of new technologies, new opportunities for their creation appear.



**Figure 1:** Component hierarchy for virtualization (a) and containerization (b) approaches

To protect such systems from malicious user code actions, the common practice is to run the code in some isolated environment. This makes the system protected from malicious solutions that may attempt to interfere with the evaluation process [5].

One potential solution to this problem can be virtual machines that emulate a hardware server. In this event, all components of a real computer are virtualized – I/O devices, hard disk, ports, etc., followed by the installation of the operating system. Although such system is isolated from the host, using virtual machines means running the entire operating system and allocating resources to emulate virtualization, which often negatively affects the performance of the platform.

An alternative way to achieve process isolation is containerization, in which the source code of the program is encapsulated along with dependencies to run on any infrastructure in an isolated environment [6]. Unlike virtual machines, containers allow you to run software in a single process, isolated from each other, on a single host, using the resources of an existing operating system. Since containers are only a part of the operating system, they are sometimes less flexible to use, for example, they can only run programs that are compatible with the current system [7].

Despite mentioned limitations, containers are widely used to create an isolated process environment. One of the most well-known examples of such containers is “chroot jail”. The idea is to copy (or create) links to the system files needed to run processes. After that, restrictions are set for the created processes – the root directory of the system is changed to the root directory of the environment. Since processes cannot refer to paths outside the modified root, they cannot perform malicious operations in these locations [8]. Containerization is the best option for the system being created, because containers allow you to run programs in an isolated environment with restrictions on system resources and at the same time work only in one process of the operating system.

### 2.3. Docker usage

One of the tools for working with containerization is Docker, which allows you to create an isolated space for a process using the Linux kernel namespace. Each container creates its own spaces with unique access, including the following:

- PID namespace (for process isolation);
- NET namespace (for managing network interfaces);
- IPC namespace (for managing access to IPC resources);
- MNT namespace (for managing file system mounting points), etc.

Docker also uses kernel control groups to allocate resources and isolate them. These groups set limits for the program for certain resources, including:

- memory control groups;
- CPU control groups;
- control groups of devices, and so on.

The containers must be used in our platform for the next purposes:

- compiling the user's code;
- running the user's code.

Therefore, it is necessary to have two containers for each programming language in the system – one container compiles the code, the second executes it. Creating containers is achieved by running images, which in turn encapsulate all the components and dependencies needed to run the application. Images are created using Dockerfile - text files with instructions created according to the rules of a Dockerfile syntax.

For example, let's take a DockerfileBase file for creating images for the Python programming language:

```
FROM python:3.7.5-slim
COPY compile_py.py .
```

This file describes the image that will be based on the python:3.7.5-slim image, the official Python version 3.7.5 image, the slim build. Copy the script to the virtual address space compile.py, which is responsible for compiling the user's code. The contents of the file are shown below:

```
import pathlib
import py_compile
import sys
```

```

if len(sys.argv) != 3:
    raise ValueError('Wrong amount of arguments!')
_, input_file, output_file = sys.argv
input_file_path = pathlib.Path(input_file)
if not input_file_path.exists():
    raise ValueError('There is no user file!')
py_compile.compile(input_file, output_file, doraise=True)

```

The base image (python-base) is created before the platform starts working. Based on this image, we create two new images – DockerfileCompiler and DockerfileRunner for compiling and running code respectively. The content of DockerfileCompiler is shown below:

```

FROM python-base
COPY ./code /code
CMD ["python", "compile_py.py", "./code/main.py", "./code-compiled/main.pyc"]

```

The content of the DockerfileRunner file is as follows:

```

FROM python-base
COPY ./code-compiled /code-compiled
CMD ["python", "/code-compiled/main.pyc"]

```

For each new user solution, a pair of such images will be created. An image based on DockerfileCompiler copies the code directory to its virtual address space and runs the script compile.py, which compiles the source code to byte-code. Due to the presence of volume (which must be configured before starting the container), the contents of the directory will remain available to other processes, in particular for the image created from the DockerfileRunner file, on the basis of which the container will execute the compiled code.

To simplify interaction with the Docker Engine API, which is an interface in the form of an HTTP client for other applications, a facade is implemented that is a wrapper over the client. Its main task is to define the interface of a real client by transferring the necessary methods to a single class:

- build-creates an image based on a specific Docker file and assigns it a name;
- run-starts a container with a specific configuration based on the specified image and returns the result of its operation;
- remove-deletes a specific image.

This class is used in “workers” – components that encapsulate the logic of working with containers and images for a given set of programming languages. Workers are waiting for the user’s solution, which is used to create a context for running code, namely:

- creating a directory with further configuring volumes;

- building a “Compiler” – an image and creating a one-time container based on it;
- building a “Runner” image based on a compiled solution.

To avoid creating new “Runner” images to run the solution with different tests, input data is passed in arguments to run the container. After testing the solution, the docker context is destroyed (images and containers are deleted).

After implementing a software component to run user solutions, the next step is to create a system that meets business requirements.

## 2.4. System’s architecture

The client-server architecture was chosen as the basis for the created software, so the selection of frameworks for developing both the backend and the frontend turned out to be an important issue.

The Python programming language (specifically version 3.7.4) was used to develop the backend. Although Django and Flask are the leading frameworks for creating Python web applications, a relatively new aiohttp framework was chosen to develop our system. The framework uses new features of the language, namely support for asynchronous programming at the syntax level [9]. The main concept of this paradigm is the use of deferred calls, which violate the usual order of instructions execution. At the same time, the resource usage model significantly differs from the model using processes and threads – there is only one thread in an asynchronous application, and new requests are served using asynchronous I/O primitives of the operating system – as a result, a small amount of resources is used for allocating new connection. Although the framework was created relatively recently (the first official “stable” release dates back to September 16, 2016), 3 major updates were released in this short period of time, and the framework repository on GitHub became the “main” one in AIO-libs, a community of programmers who develop Python packages and modules for asynchronous programming [10].

The choice of such a framework affected the choice of database, because there is only one adapter written for Python that supports asynchronous database access - that is psycopg2, a connector for PostgreSQL. Based on this adapter, the aiopg library package was created, which implements a DBAPI interface with asynchronous syntax support and is responsible for supporting SQLAlchemy, a library for working with relational databases using ORM Technology [11].

It is worth noting that not only PostgreSQL implements support for asynchronous operations. The MySQL DBMS also has it, but there is no official client for the Python programming language that would support this mode, so the PostgreSQL was chosen [12].

The React framework was chosen for the frontend. The programming code created using this framework is modular, i.e. it consists of loosely coupled components that can be reused in several places at once. Thanks to the virtual DOM, applications created using React run much faster than alternative frameworks (for example, Angular). The Redux library, a platform with unidirectional data flow and centralized data storage, was used to manage the application status. These features of the platform allow you to implement a predictable and easy-to-understand system for moving an application from one state to another.

REST was chosen as the architectural style of interaction between the backend and the frontend. Some features of this architecture are described below [13]:

- client-server architecture – separating the interface from business logic simplifies the portability of the user interface to different platforms and increases scalability by simplifying server components;
- state absence – each request to the server must contain all the necessary information for its successful execution, i.e. the server does not store any context between different requests (for example, in the form of sessions), and all information needed for the user’s “session” is stored on the client’s side (for example, an access token);
- uniform interface – following REST involves creating a “virtual” resource system, in which each of the resources has universal methods for working with it, including creating, reading, updating, and deleting. However, the true location of the entity (for example, in a database) may be located in a different virtual space.

The functional diagram shows a graph with transitions from one state of the program to another. When initializing the system, connections to the Docker Engine API are created; connecting to the database and starting the web server. After that, the system waits for a request from the user.

Each request to the system’s API is accompanied by user authorization – it checks whether the user can have access to the specified resource. If authorization fails, the user is sent a message describing the cause of the error. In other case, the request goes further through the system.

There are two main types of API requests:

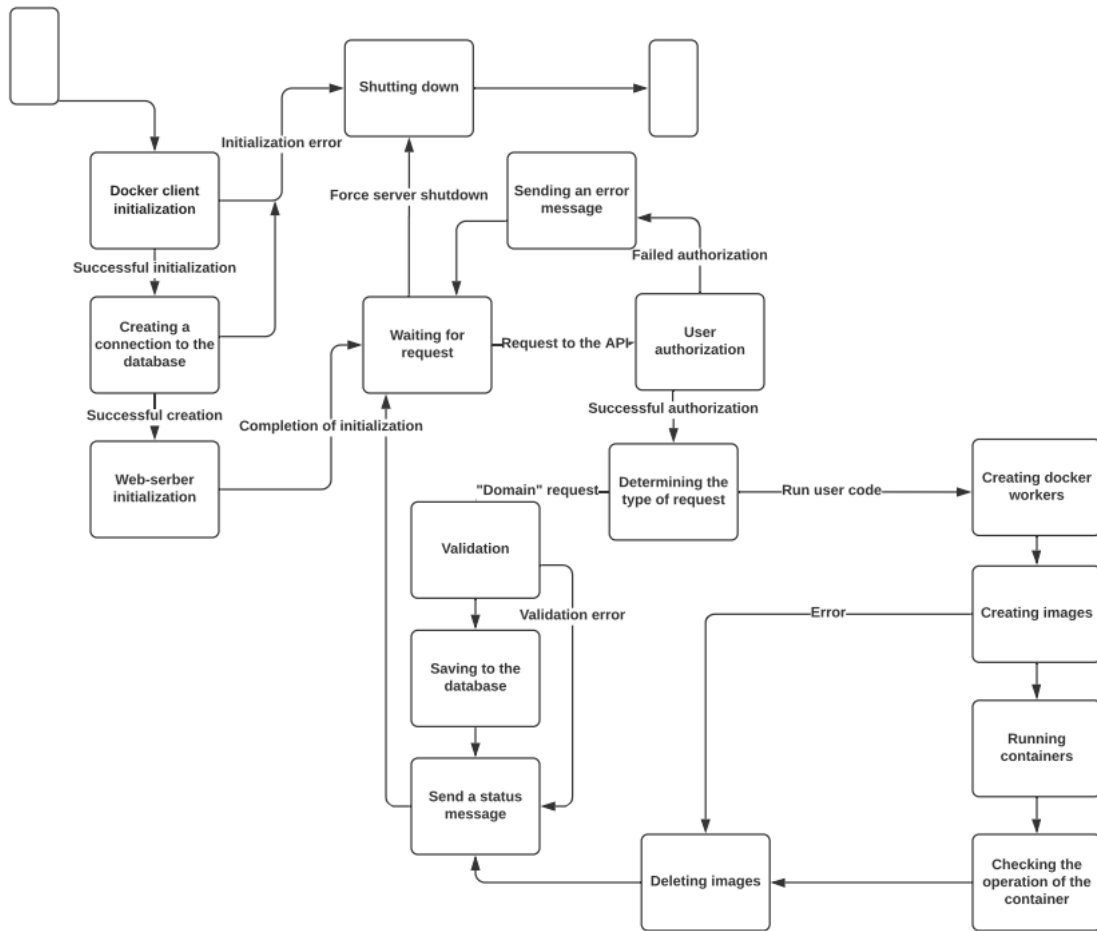
- a “CRUD request” is a request for manipulating an entity that requires interaction with the database;
- a code execution request is a request to execute code from the user in order to get the result of the program. The request requires interaction with the Docker Engine API.

## 2.5. User interface and user experience

One of the necessary requirements for the system is the speed and convenience of creating and conducting competitions, so certain restrictions were set on the system interface for effective user experience.

The most important of them is that the user should not be focusing on the elements that are not related to the context of the current resource - it is necessary to create a virtual space that does not allow the user to move randomly around the system, but on the contrary, suggests the right actions to achieve the goal.

For example, the pages “view available competitions” and “create a new competition” belong to the same type of system resources - “competitions”, so manipulations on this resource are “close” to each other in the virtual space of the system.



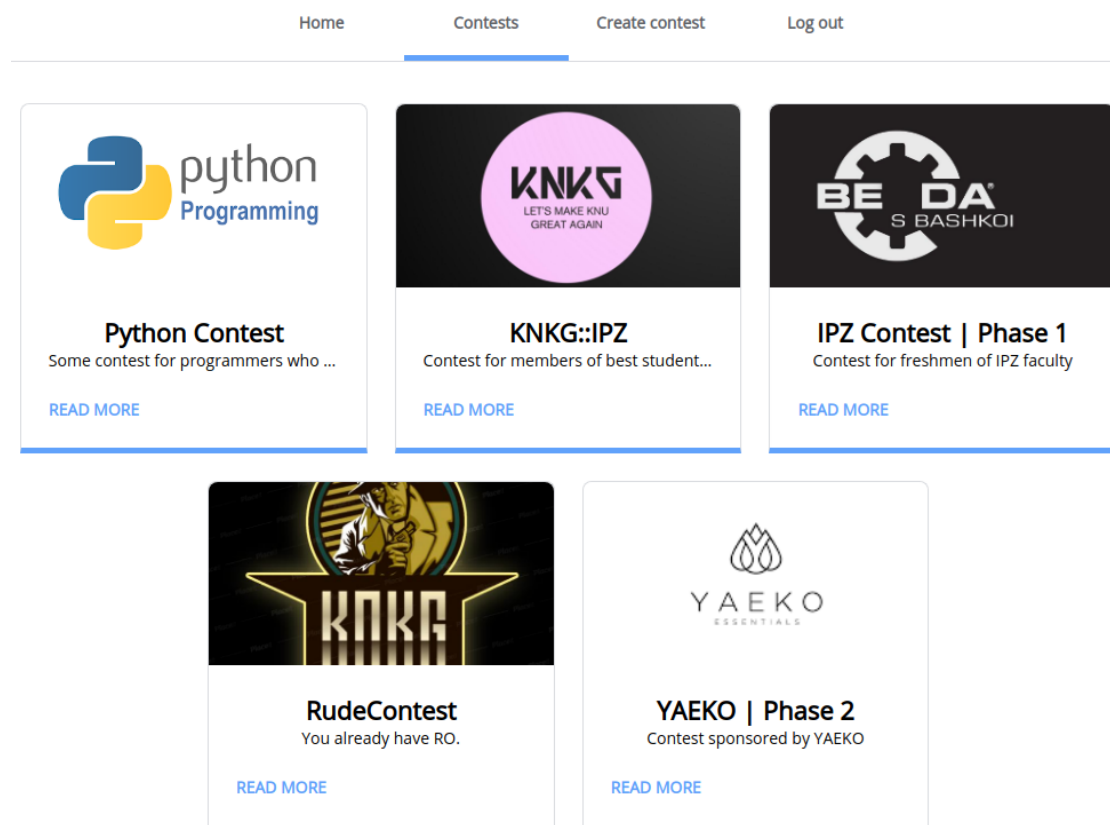
**Figure 2:** Functional diagram of the system

When viewing a specific competition, the user has more features, so the header has significantly more elements – the virtual space has expanded significantly at this stage, and links to loosely linked pages (creating a new competition) are not displayed.

The user experience of the system is based on building clear sequences of actions that lead to resources in one most intuitive way. This design allows you to focus the user’s attention on performing the planned action, instead of confusing them with oversaturated pages and long transitions.

### 3. Conclusions

Research on technologies for creating isolated environments has shown that containerization, despite a number of disadvantages, is well suited for running processes isolated from each other. The use of virtual machines, on the contrary, could lead to irrational use of system resources, so it was decided to use containers as components for running user code.



**Figure 3:** Contest list view page

After the development of this component, a system for holding contests was created, which can be used by both participants and contests organizers. A scheme for a relational database is designed for the system. The software is designed to emulate the client/server architecture, using Python and JavaScript. The system has passed several tests, which have shown an effective system operation and a great potential for further development.

## References

- [1] O. M. Markova, S. O. Semerikov, A. M. Striuk, The cloud technologies of learning: Origin, Information Technologies and Learning Tools 46 (2015) 29–44. URL: <https://journal.iitta.gov.ua/index.php/itlt/article/view/1234>. doi:10.33407/itlt.v46i2.1234.
- [2] D. Irtegov, T. Nesterenko, T. Churina, Systems for automated evaluation of programming tasks: Development, use and perspectives, Vestnik NSU. Series: Information Technologies 17 (2019) 61–73. doi:10.25205/1818-7900-2019-17-2-61-7.
- [3] D. Irtegov, T. Nesterenko, T. Churina, Development of automated evaluation systems for programming tasks, System Informatics Journal (2017) 91–116. doi:10.31144/si.2307-6410.2017.n11.



Home      Contests      **Create contest**      Log out

---

**Contest name \***


**Description \***

**Max teams \***

**Max participants in a team \***

**Date of beginning \***

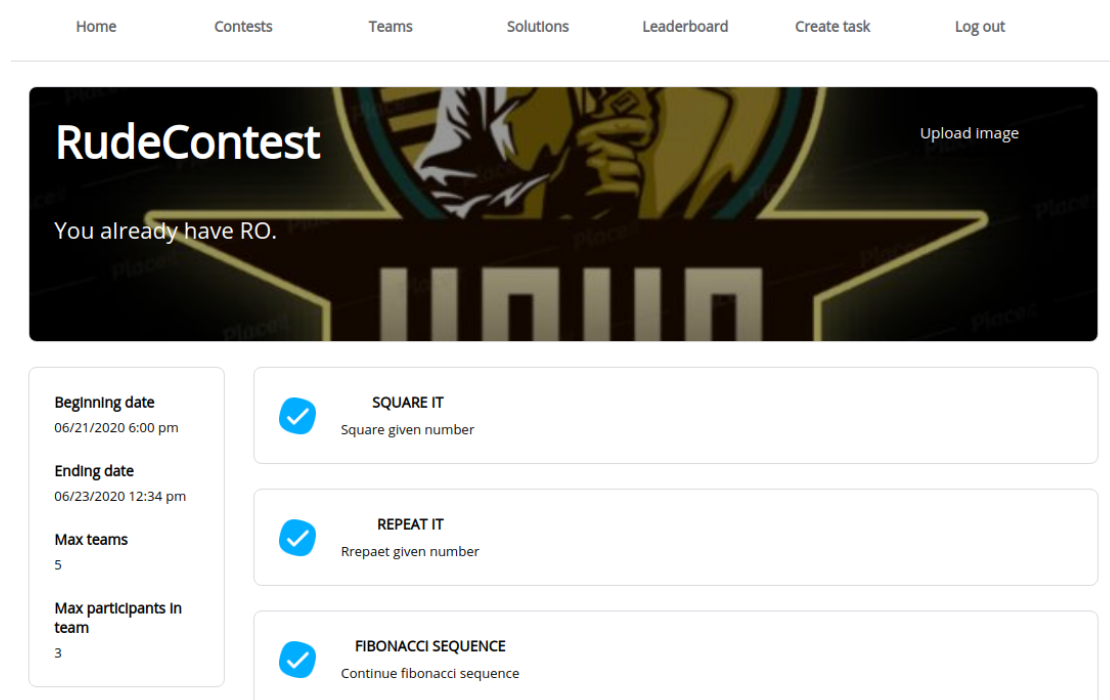
**Date of ending \***



Max file size: 5mb, accepted: jpg|gif|png

**Figure 4:** Contest creation page

- [4] ejudge system, 2020. URL: [https://ejudge.ru/wiki/index.php/%D0%A1%D0%B8%D1%81%D1%82%D0%B5%D0%BC%D0%B0\\_ejudge](https://ejudge.ru/wiki/index.php/%D0%A1%D0%B8%D1%81%D1%82%D0%B5%D0%BC%D0%B0_ejudge).
- [5] Selected papers of the International Conference joint with the XXII International Olympiad in Informatics Waterloo, Canada, August 14–21, 2010, volume 4, Institute of Mathematics and Informatics, Vilnius, Lithuania, 2010. URL: <https://ioinformatics.org/files/volume4.pdf>.
- [6] IBM Cloud Education, Containerization explained, 2019. URL: <https://www.ibm.com/cloud/learn/containerization>.
- [7] J. Turnbull, The Docker Book: Containerization Is the New Virtualization, 2014. URL: <https://dockerbook.com/>.
- [8] chroot “jail” - what is it and how do I use it?, 2010. URL: <https://unix.stackexchange.com/questions/105/chroot-jail-what-is-it-and-how-do-i-use-it>.
- [9] Y. Selivanov, PEP 492 – Coroutines with async and await syntax, 2015. URL: <https://www>.



**Figure 5:** Contest resources overview page

- [python.org/dev/peps/pep-0492/](https://python.org/dev/peps/pep-0492/).
- [10] aio-lib: The set of asyncio-based libraries built with high quality, 2021. URL: <https://github.com/aio-lib>.
  - [11] A. Svetlov, A. Firsov, Welcome to AIOPG, 2020. URL: <https://aiopg.readthedocs.io/en/stable/>.
  - [12] Is there an asynchronous Python client for MySQL with pooling features?, 2015. URL: <https://www.quora.com/Is-there-an-asynchronous-Python-client-for-MySQL-with-pooling-features>.
  - [13] restfulapi.net, What is REST, 2020. URL: <https://restfulapi.net/>.

# Development of software and hardware complex of GPS-tracking

Mykola V. Klymenko, Andrii M. Striuk

*Kryvyi Rih National University, 11 Vitalii Matusevych Str., Kryvyi Rih, 50027, Ukraine*

## Abstract

The paper considers the typical technical features of GPS-tracking systems and their development, as well as an analysis of existing solutions to the problem. Mathematical models for the operation of hardware and software of this complex have been created. An adaptive user interface has been developed that allows you to use this complex from a smartphone or personal computer. Methods for displaying the distance traveled by a moving object on an electronic map have been developed. Atmega162-16PU microcontroller software for GSM module and GPS receiver control has been developed. A method of data transfer from a GPS tracker to a web server has been developed. Two valid experimental samples of GPS-trackers were made and tested in uncertain conditions. The GPS-tracking software and hardware can be used to monitor the movement of moving objects that are within the coverage of GSM cellular networks.

## Keywords

satellite navigation system, GPS receiver GN-801, microcontroller Atmega162-16pu, transport monitoring, GPS, GSM, GPRS, GPS tracker, SIM800L, Atmel Studio 7

## 1. Introduction

Today, most companies of all forms of ownership use vehicles to carry out their business activities. The number of such vehicles can number several hundred in one enterprise, for example, fleets of police units, municipal road cleaning and repair equipment, ambulances, taxis, car and construction equipment rental agencies, delivery services, passenger transport, fire departments, agricultural equipment and others. All these vehicles are movable property that requires operational management and supervision by the head of the enterprise for the successful implementation of economic activities.

It is important for the head of the company to know the location of the vehicle at any time to control the work schedule or delivery of goods. Also necessary information is the history of movement of the vehicle for some period of time to prevent misuse of the company's fleet, which leads to additional costs.

However, not only business leaders but also anyone wants to control the location of their movable property. Common car thefts require prompt search. In such situations, every second


---

*CS&SE@SW 2020: 3rd Workshop for Young Scientists in Computer Science & Software Engineering, November 27, 2020, Kryvyi Rih, Ukraine*

✉ [fxnik777@gmail.com](mailto:fxnik777@gmail.com) (M.V. Klymenko); [andrey.n.stryuk@gmail.com](mailto:andrey.n.stryuk@gmail.com) (A.M. Striuk)

🌐 <http://mpz.knu.edu.ua/pro-kafedru/vikladachi/224-andrii-striuk> (A.M. Striuk)

🆔 0000-0001-9240-1976 (A.M. Striuk)

© 2020 Copyright for this paper by its authors.  
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).  
 CEUR Workshop Proceedings (CEUR-WS.org)

is decisive.

It is possible to help solve these problems by developing GPS-tracking software and hardware systems that use data obtained from satellite navigation systems, such as GPS (Global Positioning System) or GLONASS (Global Navigation Satellite System).

GPS-tracking software and hardware can be used in various sectors of the economy to monitor a variety of mobile objects in real time, which are within the coverage of GSM-cellular networks. Such objects can be not only vehicles, but also employees of the enterprise, a group of tourists in an unfamiliar country, children, animals, mobile equipment.

To achieve the goal, you must perform the following tasks:

- to analyze the current application of GPS-tracking;
- analyze the existing means of GPS-tracking;
- develop a functional diagram of the software and hardware complex of GPS-tracking;
- develop mathematical software for this complex;
- develop an electrical circuit of the hardware of this complex;
- develop the structure of the database;
- develop a user interface for the software and hardware complex;
- implement an test the hardware and software complex.

## **2. Urgency of tasks of monitoring of mobile objects and means of their solution**

### **2.1. Analysis of the modern application of GPS-tracking in industry and society**

A typical GPS tracking system consists of three structural elements:

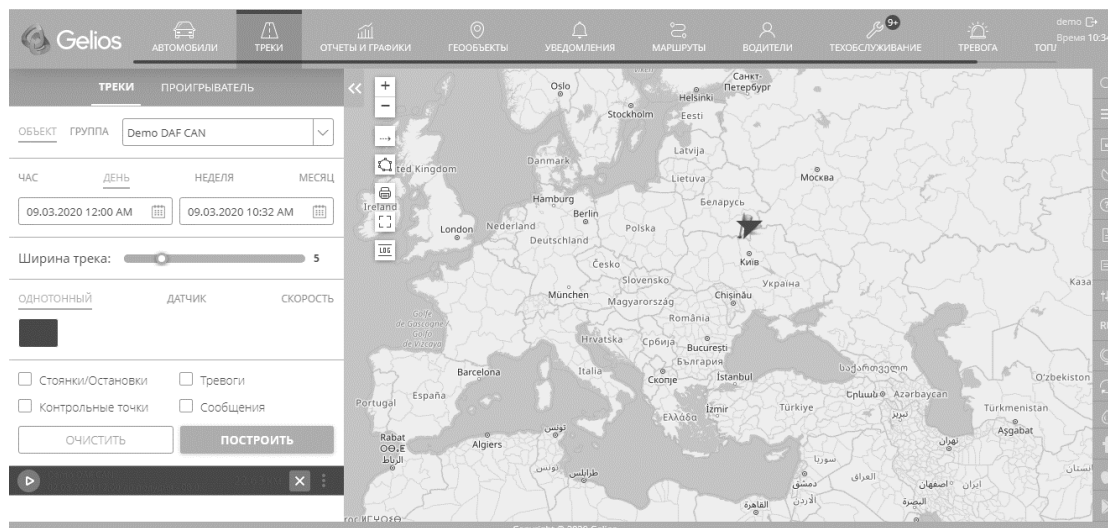
1. Terminals that are installed on the transport.
2. A server that stores information about the movement of the vehicle.
3. Client workplaces, with appropriate software for transport monitoring.

The terminal is a specialized GPS tracker that contains a GPS receiver and cellular tools.

The server can be either a regular personal computer with server software installed for simple GPS tracking systems, or a large distributed system with specialized software for complex specialized GPS tracking systems.

In some cases, the client software can be combined into one program together with the server part. However, it is often permissible to connect many workstations to the server. In some GPS tracking systems, thanks to the installation of special software on client computers, it is possible to quickly obtain information through the use of web feeds.

Therefore, the operation of the GPS-tracking system is divided into the following stages:



**Figure 1:** “Gelios” user interface

- GPS-tracker calculates the geographical coordinates of the vehicle;
- GPS-tracker transmits the received coordinates and other data to the server via cellular or satellite communication channels;
- the server receives GPS tracker data, analyzes and writes them to the database;
- the customer receives information about moving the vehicle to a smartphone or personal computer.

## 2.2. Analysis of existing GPS tracking tools

Today, a large number of hardware and software solutions are offered for vehicle monitoring on the GPS tracking system market.

Figure 1 shows the user interface of the software-hardware complex of GPS-tracking “Gelios”, developed by the company “GeliosSoft” ([www.geliossoft.ru](http://www.geliossoft.ru)). Users access the system through a web browser.

Real-time transport monitoring with the Gelios system makes it possible to increase the efficiency of vehicle operation, optimize money costs, increase the level of transport safety, and improve the work discipline of the company’s staff. Gelios GPS-tracking software and hardware uses third-party GPS trackers, one of which is shown in figure 2.

This tracker is designed to determine the exact location of the object and other parameters and then transfer them to the server via GSM networks. The device has two analog and four digital inputs, which allows you to connect fuel level sensors, axle pressure sensor, temperature sensors and more.



Figure 2: “Ruptela” GPS-tracker

### 2.3. Principles of satellite navigation

The main parts of satellite navigation systems [1] are:

- space equipment;
- ground equipment;
- custom equipment.

Space equipment includes navigation satellites. The main functions of navigation satellites include the transmission of the necessary information as part of its radio signal to determine the location of ground objects. The ground equipment includes a control center, spaceport, measuring equipment. The control center coordinates the work of all parts of the satellite navigation system. The user segment includes satellite receivers of consumers. They are designed to receive signals from navigation satellites, calculate navigation parameters and process calculations.

### 2.4. Basics of GSM cellular networks

Cellular communication is a type of mobile radio communication based on a cellular network [2]. The main feature of cellular communication is that the coverage area is divided into separate cells, which is due to the service areas of the base stations of the BS. Zones of cellular cells intersect in some places and together form a cellular network. The service area of the base station on a flat surface, in its form, is a circle. The cellular network composed of these circles has the form of hexagons.

The network consists of cellular base stations spaced apart. They allow you to determine the location of moving objects, but with a significant error, which can reach several hundred

meters. Base stations provide continuity of communication as a network subscriber moves from one cell area to another.

When a cell phone leaves the coverage area of a base station with a radius of approximately 35 kilometers, the phone establishes a connection with the base station using the base station. The phone measures the signal level of the thirty-two nearest base stations simultaneously. Data on the top six signal levels are transmitted by telephone through the service channel to the switching center. Then, the communication control is transferred from one station to another. The maximum distance between the phone and the nearest base station for communication is 120 kilometers, but with the use of GSM amplifiers and directional antennas.

Base stations of the cellular network are located in the center of the cell on top of the iron towers, on the roofs of multi-storey residential buildings. The radio is constantly scanned by the cell phone to search for the base station signal. If a base station signal is found, the phone transmits an ID to that station. The telephone and the base station are in constant radio contact. Different cellular networks are connected to each other and to a fixed telephone network.

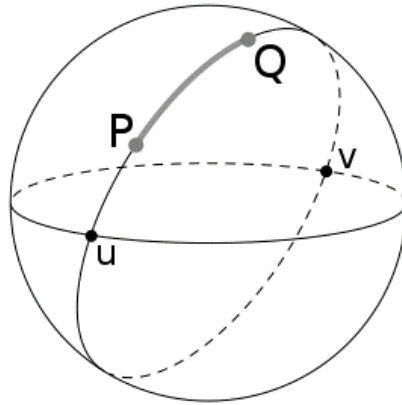
GSM cellular networks are 2G networks, i.e. the second generation of digital cellular communication. Mobile phones are made to support the ability to operate in four frequency bands: 850, 900, 1800, 1900 MHz, because different parts of the world use different frequency bands for GSM networks. GSM-network allows you to use a common packet data service, i.e. GPRS. This is a standard that uses a voice line to transmit any information. The service is used by mobile devices to access the Internet. There are GPRS classes from the first to the twelfth. As the class grows, the data rate increases. The average GPRS data rate is 20 to 40 kilobits per second.

### **3. Design of software and hardware complex of GPS-tracking**

#### **3.1. Design of the functional scheme of the software and hardware complex of GPS-tracking**

The work of the software and hardware complex of GPS-tracking should consist of the following stages:

- GPS-tracker receiver calculates the geographical coordinates of its position;
- the microcontroller of the tracker receives geographical coordinates from the GPS receiver;
- GSM-modem of the tracker receives geographical coordinates from the microcontroller and transmits them to the web server via GSM-network;
- the web server receives the geographical coordinates of the GPS-receiver, checks them for integrity and stores them in the database;
- the user software displays the received new geographical coordinates of the tracker on the electronic map in the web browser window.



**Figure 3:** Two points on the sphere

Let's make the functional scheme of this complex.

You must have two domains on the web server. The domain mygps.pp.ua must be installed SSL certificate [3] to use the HTTPS protocol [4] when exchanging data between the user and the web server. The second domain will be used to receive geographical coordinates from the tracker via HTTP protocol [4] using PHP scripts [5]. Trackers must transmit geographical coordinates in the parameters of the GET request [5]. The two domains must use a common database. It has been decided to develop two types of GPS trackers. The first type of trackers will transmit the coordinates of its position approximately every 15 seconds, working continuously. The data rate of such a tracker will allow you to display the movement of the car with sufficient accuracy.

The second type of trackers will transmit the coordinates of their position approximately every 5 minutes. After each successful transfer of coordinates, the tracker will enter sleep mode, which is designed to save battery power, which will significantly increase the battery life of the device.

Each tracker will have a unique number for the tracker to access the web server. To access the Internet, the Internet tracker will use sim-cards of the cellular operator "lifecell".

### 3.2. Development of mathematical model for GPS-tracking complex

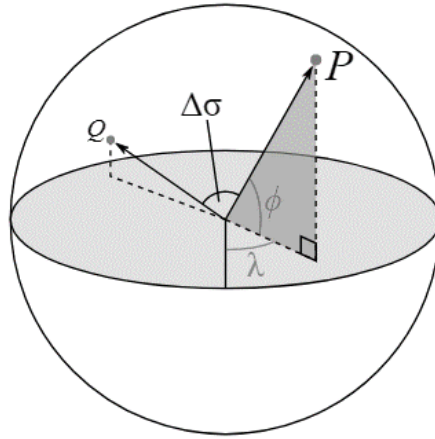
Consider a mathematical model for calculating the distance between two points on the surface of the globe by geographical coordinates. The shape of the Earth can be described as a sphere. In Euclidean geometry, the smallest distance between two points is a straight line. However, straight lines are not possible in the sphere. These lines on the sphere are part of large circles [6], the centers of which coincide with the center of the sphere (figure 3).

Figure 4 shows a sphere as a model of the globe. On the surface of the sphere are points  $Q$ ,  $P$ , the distance between which must be calculated.

Therefore, the distance  $L$  between points  $Q$  and  $P$  is calculated by the formula:

$$L = \Delta\sigma R,$$





**Figure 4:** Model of the globe with two points on the surface

where  $\Delta\sigma$  is the angular distance between the points  $Q$  and  $P$ , rad;  $R$  - radius of the Earth, m.

The angular distance  $\Delta\sigma$  between the points  $Q$  and  $P$  is calculated by the modified formula of haversine [6, 7]:

$$\Delta\sigma = \arctan \frac{\sqrt{(\cos \phi_2 \sin(\lambda_2 - \lambda_1))^2 + (\cos \phi_1 \sin \phi_2 - \sin \phi_1 \cos \phi_2 \cos(\lambda_2 - \lambda_1))^2}}{\sin \phi_1 \sin \phi_2 + \cos \phi_1 \cos \phi_2 \cos(\lambda_2 - \lambda_1)},$$

where  $\phi_1, \phi_2$  - latitudes of two points, rad;  $\lambda_1, \lambda_2$  - longitudes of two points.

Consider a mathematical model for calculating the data rate for the USART microcontroller Atmega162-16PU. USART (Universal Synchronous and Asynchronous serial Receiver and Transmitter) is a universal synchronous-asynchronous serial transceiver. This is a peripheral device that is part of the microcontroller Atmega162-16PU [8] from the company "Microchip". USART data rate is the number of bits transmitted or received per second.

The data transfer rate via USART is calculated by the formula:

$$BAUD = \frac{f_{osc}}{16(UBRR0 + 1)},$$

where  $f_{osc}$  is the clock frequency of the microcontroller, Hz;  $UBRR0$  - the contents of the register pair UBRR0 in decimal form.

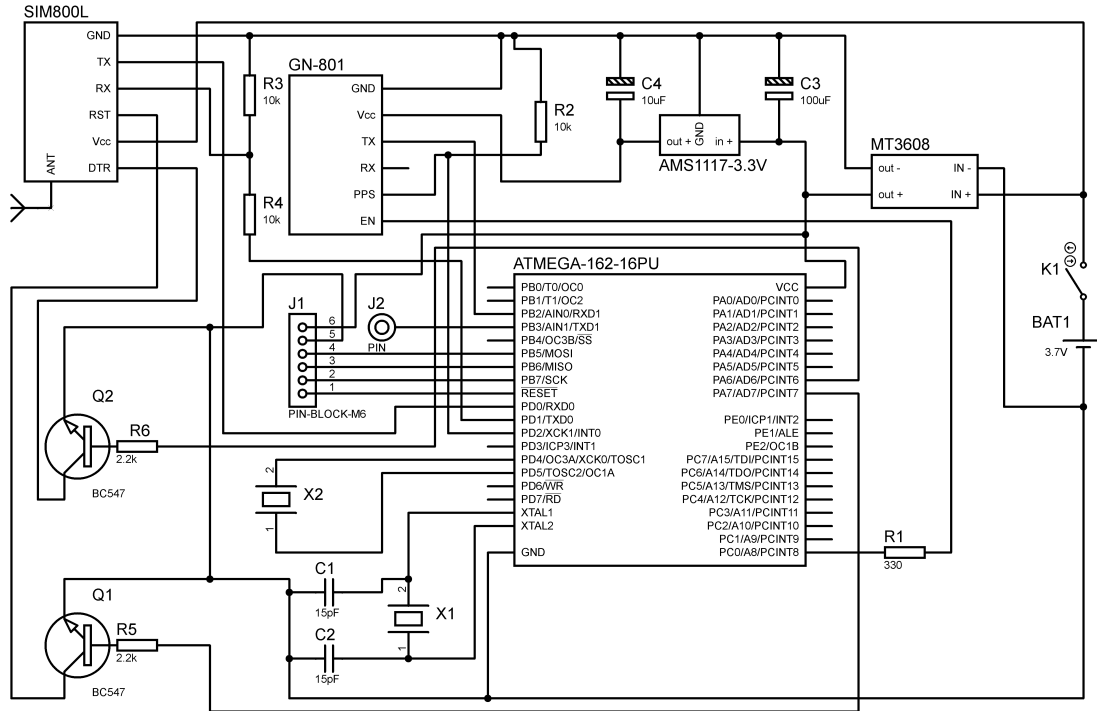
Consider a mathematical model for calculating the overflow time of 16-bit timers-counters of the microcontroller Atmega162-16PU. Numbers from 0 to 216 - 1 can be written in 16-bit case.

Overflow time of 16-bit timers-counters is calculated by the formula:

$$t = T(2^{16} - 1),$$

where  $T$  is the increment time of the 16-bit register.

The increment time of the 16-bit register is calculated by the formula:



**Figure 5:** Electronic circuit of the GPS tracker

$$T = \frac{d}{f_{osc}},$$

where  $d$  is the clock divider of the microcontroller;  $f_{osc}$  - clock speed of the microcontroller, Hz.

### 3.3. Development of electrical circuit diagram of hardware

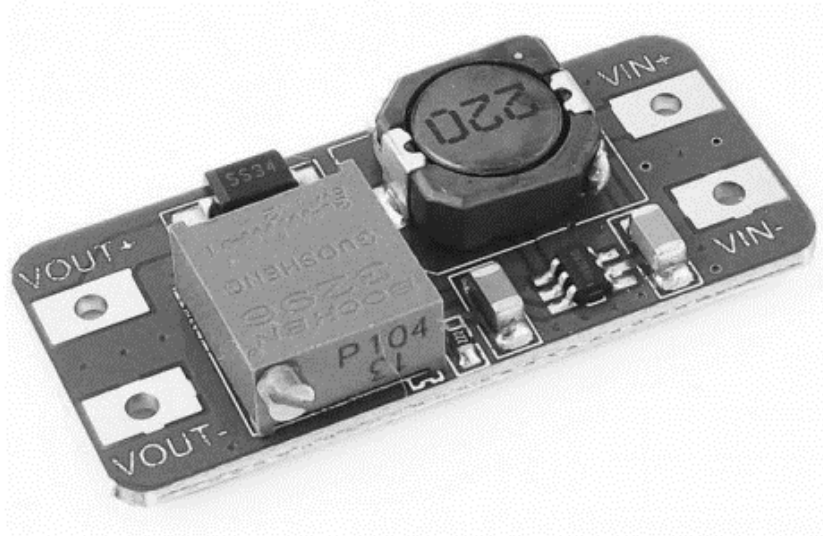
The development of the electrical circuit of the GPS-tracker, which is shown in figure 5, was carried out using a specialized software package "Proteus 8 Professional".

Consider the design features of this scheme. The power supply of the tracker is a lithium-ion battery with a voltage of 3.7 volts and a built-in short circuit protection circuit. The tracker is turned on and off by switch K1.

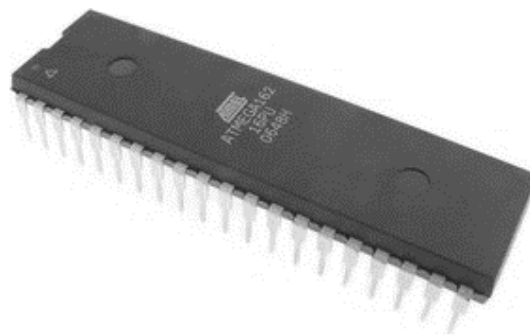
The MT3608 module [9] is designed to increase the voltage from 3.7 to 5 volts. This module is shown in figure 6.

The AMS1117-3.3 circuit is designed to reduce the voltage from 5 to 3.3 volts. Polar electrolytic capacitors, denoted as C3, C4, have a capacity of  $100 \mu F$  and  $10 \mu F$ , respectively.

The Atmega162-16PU microcontroller [8], shown in figure 7, is powered by the MT3608 module.



**Figure 6:** Appearance of the MT3608 module



**Figure 7:** Atmega162-16PU microcontroller

This microcontroller is programmed through the contact group J1 by USBASP programmer [10], which is shown in figure 8.

Quartz resonators X1, X2 have frequencies of 4 MHz and 32768 Hz, respectively. Resonator X1 is designed to clock the microcontroller. Resonator X2 in the microcontroller helps to implement time measurements. According to the [8], the capacitances of the capacitors C1, C2 can be from 12 to 22 pF. The GPS receiver GN-801 [11, 12] is shown on two sides in figure 9.

This receiver has a UART interface with RX, TX contacts. The RX contact is for receiving data and the TX is for transmitting. The PPS contact with its high signal level warns the microcontroller about data transmission. The presence or absence of voltage at the EN contact turns the GPS receiver on or off. To activate the EN contact, it is necessary to remove the resistor from the surface of the GPS receiver as shown in the [11].

Resistors R3, R4 are a voltage divider and have a resistance of 10 k ohm each.



**Figure 8:** USBASP programmer



**Figure 9:** GN-801 GPS receiver

The GSM module SIM800L [13], which is shown in figure 10, is designed to transmit data to a web server via the Internet. The module uses micro-SIM phone cards.

This GSM module has a UART interface [8] with RX, TX contacts. The RST contact is for rebooting the device. The DTR contact is designed to exit the GSM module from sleep mode. The ANT contact is for connecting a GSM antenna.

Bipolar transistors BC547, denoted in the diagram as Q1 and Q2, are designed to reboot the GSM-module and wake it from sleep, respectively. When you open the transistor Q1, the voltage level at the contact RST changes from high to low. When you open the transistor Q2,



**Figure 10:** GSM module SIM800L

the voltage level at the contact DTR changes from high to low. Resistors R5, R6 are designed to limit the current through the bases of transistors Q1, Q2. Each of these resistors has a resistance of 2.2 k ohm.

## 4. Results

### 4.1. Implementation of hardware

The Atmel Studio 7 development environment was chosen for the development of the Atmega162-16PU microcontroller software [14]. It is an integrated platform for developing and debugging AVR and SAM (Smart ARM-based Microcontroller) programs. Atmel Studio 7 provides an easy-to-use environment for writing, compiling, and debugging programs written in C/C++ or assembly code. This environment supports interaction with debuggers, programmers, development kits that support AVR or SAM devices.

Consider data transfer to a web server. The process of data transfer to the web server is based on the cyclic execution of a constant sequence of AT-commands [15] by the GSM-module, the list of which is shown in figure 11.

The command in line 62, i.e. "AT", checks the readiness of the module to execute commands. The command in line 63 checks the registration of the module in the cellular network. Commands in lines 64, 65, 66 connect to GPRS. The command in line 67 initializes the HTTP session. The command in line 68 sets the ID for the HTTP session. The command in line 69 sets the GET request. The "id" parameter is assigned the value of the tracker number by which the tracker is identified on the web server. If a tracker has a number that is not in the database on the web server, the data of such a tracker is not stored in the database. The gps-data parameter is assigned the value of the data packet from the GPS receiver. The command in line 70 sets the header "User-agent" of the HTTP request. The command in line 71 executes a GET request. The command in line 72 terminates the HTTP session. The command on line 73 termi-

```

62 char* at[] = {"AT\r",
63             "AT+CCALR?\r",
64             "AT+SAPBR=3,1,\"CTYPE\","GPRS"\r",
65             "AT+SAPBR=3,1,\"APN\","internet"\r",
66             "AT+SAPBR=1,1\r",
67             "AT+HTTTPINIT\r",
68             "AT+HTTTPARA=\"CID\",1\r",
69             "AT+HTTTPARA=\"URL\", \"http://fxnik7fl.beget.tech/php/addGps.php?id=1234567&gpsdata=
70             "AT+HTTTPARA=\"UA\", \"Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML,
71             "AT+HTTTPACTION=0\r",
72             "AT+HTTPTERM\r",
73             "AT+SAPBR=0,1\r",
74             "AT+CSCLK=1\r",
75             "AT+CSCLK=0\r"};

```

**Figure 11:** GSM module SIM800L

nates the GPRS connection. The command in line 74 enters the GSM module into sleep mode. The command in line 75 wakes the GSM module from sleep mode. It is not enough to simply send the last command to the GSM module, as in this mode the UART interface of this module is blocked. To unlock the UART of this module, you need to change the voltage level from high to low at the DTR contact. Next, execute the command in line 75. After executing each command, the GSM module returns an alphanumeric response code that must be received and analyzed. Commands and response codes are exchanged through the USART0 interface of the microcontroller.

As a result of hardware development for the GPS-tracking complex, two working prototypes of GPS-trackers are made, which are shown in figure 12.

These GPS trackers are the same in structure, however, each of them may have different software.

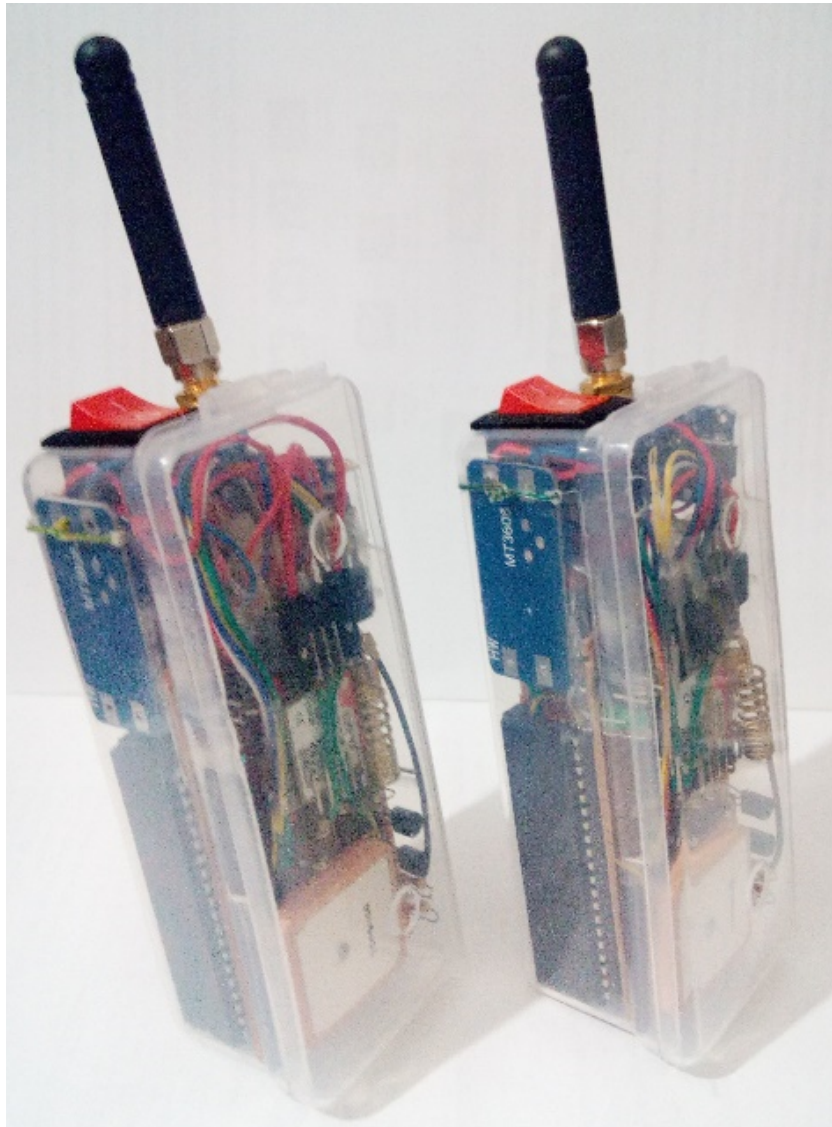
## 4.2. Testing of software and hardware complex of GPS-tracking

As a result of testing the software and hardware complex of GPS-tracking, the movement of the vehicle in the city of Kryvyi Rih on Kuprina and Petro Kalnyshevsky streets was traced, which is shown in figure 13.

In this figure, the path of movement is shown by black circles, which are connected by a gray line. If you click on such a circle, a pop-up window with additional information appears. Among this information is the length of the track from the beginning to this point.

## 5. Conclusions

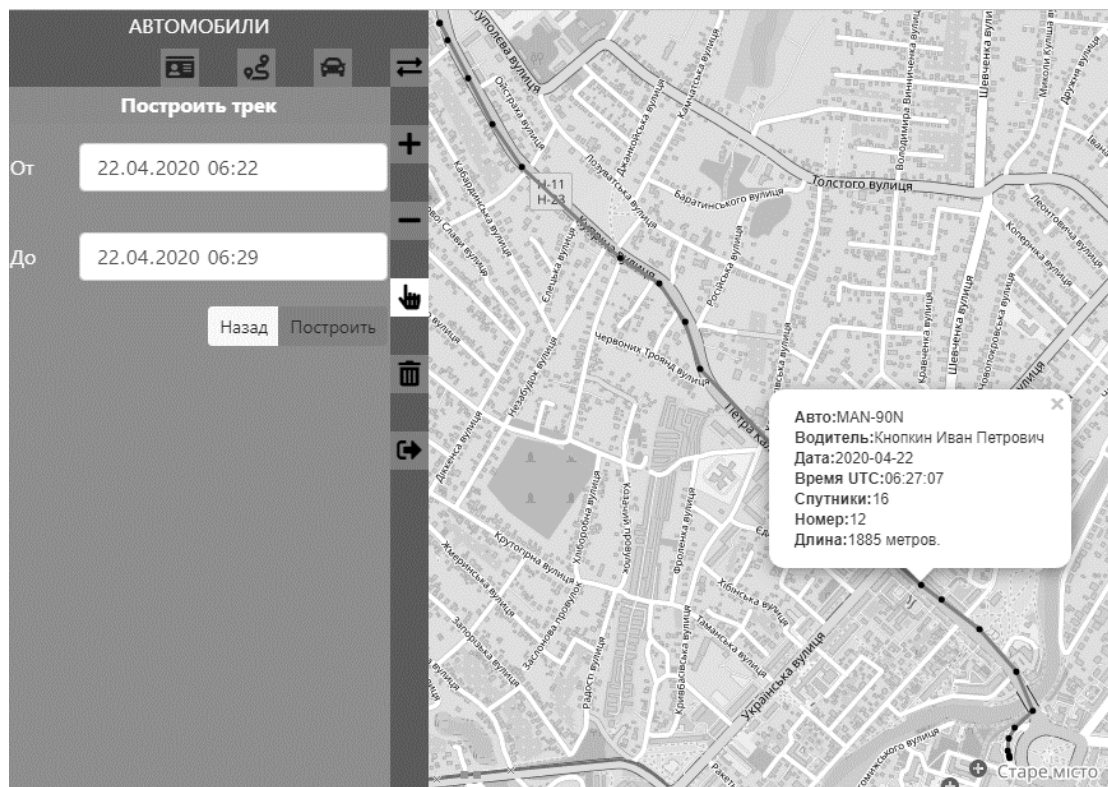
Today, satellite monitoring of moving objects is still relevant. The latest advances in science and technology are involved in this activity. In this work the design and implementation of software and hardware complex of GPS-tracking with the use of modern web technologies and electronic components is carried out. On the basis of the analysis of modern application and existing means of GPS-tracking, the decomposition of these systems is carried out. GPS-trackers, web server, client software are the main structural elements of software-hardware complexes of GPS-tracking. Based on the analysis, a functional diagram and mathematical



**Figure 12:** Prototypes of GPS trackers

software of GPS-tracking software and hardware have been developed. Based on the analysis of the functions of GPS-trackers, the necessary electronic components, such as microcontroller, GPS-receiver, GSM-module, were selected. It was found that they all have a serial UART data interface. Based on research data, an electrical circuit of a GPS tracker has been developed. A web-based user interface of the software and hardware complex has been developed, which allows access to this system via a computer or smartphone using popular web browsers. Based on the design results of this GPS-tracking software and hardware, software for the Atmega162-16PU microcontroller, as well as user software with a web interface have been developed.

Testing the operation of this software and hardware complex of GPS-tracking was carried



**Figure 13:** The path of movement of the vehicle

out in the form of tracking the path of movement of the vehicle in real time. Testing has shown that the accuracy of the tracker's measurements of the coordinates of a moving object depends on the environment. The fewer high-rise buildings around the tracker, the more accurate the coordinates. GPS-tracking software and hardware can be used to monitor the movement of moving objects that are within the coverage of GSM cellular networks. Such objects can be not only vehicles, but also employees of the enterprise, a group of tourists in an unfamiliar country, children, animals, mobile equipment.

## References

- [1] N. Bonnor, A brief history of global navigation satellite systems, *Journal of Navigation* 65 (2012) 1–14. doi:10.1017/S0373463311000506.
- [2] P. Morgan, Approaches to optimising radio performance of gsm/gprs/edge/w-cdma cellular handsets, *Microwave Engineering Europe* (2004) 23–26.
- [3] P. Robinson, *Understanding digital certificates and secure sockets layer (SSL)*, CRC Press, 2004. doi:10.1201/9780203498156.
- [4] D. Naylor, A. Finamore, I. Leontiadisz, Y. Grunenbergerz, M. Mellia, M. Munafò, K. Pagiannakiz, P. Steenkiste, The cost of the "s" in https, *Association for Computing Ma-*



- chinery, Inc, 2014, pp. 133–139. doi:10.1145/2674005.2674991, conference of 10th ACM International Conference on Emerging Networking Experiments and Technologies, CoNEXT 2014 ; Conference Date: 2 December 2014 Through 5 December 2014; Conference Code:125893.
- [5] W. Jason Gilmore, *Beginning PHP and MySQL 5: From novice to professional*, Apress Media LLC, 2006. doi:10.1007/978-1-4302-0117-5.
  - [6] M. Dubinin, Calculation of distance and initial azimuth between two points on the sphere, 2006. URL: <https://gis-lab.info/qa/great-circles.html>.
  - [7] C. Veness, Calculate distance, bearing and more between Latitude/Longitude points, 2020. URL: <http://www.movable-type.co.uk/scripts/latlong.html>.
  - [8] Atmel, 8-bit AVR Microcontroller with 16K Bytes In-System Programmable Flash. ATmega162, ATmega162V, 2013. URL: [https://ww1.microchip.com/downloads/en/DeviceDoc/Atmel-2513-8-bit-AVR-Microcontroller-ATmega162\\_Datasheet.pdf](https://ww1.microchip.com/downloads/en/DeviceDoc/Atmel-2513-8-bit-AVR-Microcontroller-ATmega162_Datasheet.pdf).
  - [9] Aerosemi Technology Co.,Ltd, MT3608: High Efficiency 1.2 MHz 2A Step Up Converter, 2012. URL: <https://www.mikrocontroller.net/attachment/212877/MT3608.pdf>.
  - [10] USBASP Programmer for AVR microcontrollers, 2013. URL: [https://arduino.ua/docs/USBASP\\_Datasheet.pdf](https://arduino.ua/docs/USBASP_Datasheet.pdf).
  - [11] TOPGNSS, STOTON GNSS Module. GN-801 UART GNSS Module Ubx- 8030-kt chipset, 2018. URL: <http://www.stotoncn.com/upload/file/201801/1516180733487356.pdf>.
  - [12] u-blox, u-blox 8 / u-blox M8: Receiver description. Including protocol specification, 2021. URL: [https://www.u-blox.com/sites/default/files/products/documents/u-blox8-M8\\_ReceiverDescrProtSpec\\_%28UBX-13003221%29.pdf](https://www.u-blox.com/sites/default/files/products/documents/u-blox8-M8_ReceiverDescrProtSpec_%28UBX-13003221%29.pdf).
  - [13] Downloads Simcom.ee, 2019. URL: <https://simcom.ee/documents/?dir=SIM800L>.
  - [14] Microchip Technology, Microchip Studio for AVR® and SAM Devices, 2021. URL: <https://www.microchip.com/en-us/development-tools-tools-and-software/microchip-studio-for-avr-and-sam-devices>.
  - [15] SIMCom, SIM800 Series\_AT Command Manual\_V1.09, 2015. URL: [https://www.elecrow.com/wiki/images/2/20/SIM800\\_Series\\_AT\\_Command\\_Manual\\_V1.09.pdf](https://www.elecrow.com/wiki/images/2/20/SIM800_Series_AT_Command_Manual_V1.09.pdf).