

**МАТЕМАТИЧЕСКОЕ И ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ ВЫЧИСЛИТЕЛЬНЫХ СИСТЕМ,
КОМПЛЕКСОВ И КОМПЬЮТЕРНЫХ СЕТЕЙ / MATHEMATICAL SOFTWARE FOR COMPUTERS,
COMPLEXES AND COMPUTER NETWORKS**

DOI: <https://doi.org/10.23670/IRJ.2023.130.1>

**DEVELOPMENT OF PROCEDURE FOR DIAGNOSTICS OF CARDIAC NOISE BY ANALYSIS OF
PHONOCARDIOGRAMS**

Research article

Ilichev V.Y.^{1,*}, Lavrentyeva G.V.², Kashirin D.S.³

¹ORCID : 0000-0003-1017-5544;

²ORCID : 0000-0003-2962-8427;

³ORCID : 0000-0001-8942-1925;

^{1,2,3} Bauman Moscow State Technical University, Kaluga, Russian Federation

* Corresponding author (patrol8[at]yandex.ru)

Abstract

The purpose of the described work is to create a method for determining different types of cardiac noise by analyzing phonocardiograms (PKG) of patients. In the course of research, a work plan was developed, an algorithm and code of the author's program were created in the modern high-level Python language using libraries of special functions for reading PKG, constructing their oscillograms and spectrograms, conducting two types of analysis of spectrograms and displaying results on a computer screen and in files.

Based on the results of processing the above demonstration examples of PKG from an extensive database of real records of heart work, conclusions are made about the nature of the identified physiological problems in each specific case. The conclusion outlines further steps to supplement the described methodology with automated diagnostics of heart disease using hardware and software tested earlier in other studies. The promising directions for using the developed methodology are presented for implementation by the team of authors.

Keywords: cardiac noise, phonocardiogram, oscillogram, spectrogram, fast Fourier transform, Python language.

**РАЗРАБОТКА МЕТОДИКИ ДИАГНОСТИКИ СЕРДЕЧНЫХ ШУМОВ ПУТЁМ АНАЛИЗА
ФОНОКАРДИОГРАММ**

Научная статья

Ильичев В.Ю.^{1,*}, Лаврентьева Г.В.², Каширин Д.С.³

¹ORCID : 0000-0003-1017-5544;

²ORCID : 0000-0003-2962-8427;

³ORCID : 0000-0001-8942-1925;

^{1,2,3} Московский государственный технический университет имени Н.Э. Баумана, Калуга, Российская Федерация

* Корреспондирующий автор (patrol8[at]yandex.ru)

Аннотация

Целью описываемой работы является создание методики определения разных видов сердечного шума путём анализа фонокардиограмм (ФКГ) пациентов. В ходе исследований разработан план работ, созданы алгоритм и код авторской программы на современном высокоуровневом языке Python с использованием библиотек специальных функций для чтения ФКГ, построения их осциллограмм и спектрограмм, проведения двух видов анализа спектрограмм и вывода результатов на экран компьютера и в файлы.

По результатам обработки приведённых демонстрационных примеров ФКГ из обширной базы данных по реальным записям работы сердца делаются заключения о характере выявленных физиологических неполадок в каждом конкретном случае. В заключении изложены дальнейшие шаги по дополнению описанной методики средствами автоматизированной диагностики сердечных заболеваний с помощью аппаратных и программных средств, опробованных ранее в других исследованиях. Приводятся планируемые для реализации коллективом авторов перспективные направления использования разработанной методики.

Ключевые слова: сердечный шум, фонокардиограмма, осциллограмма, спектрограмма, быстрое преобразование Фурье, язык Python.

Introduction

In modern medicine, automated diagnostic systems play an invaluable role in making a correct diagnosis to patients [1]. Their use becomes necessary in connection with the rapid increase in the flow of patients in medical institutions, the emergence of new diseases and varieties of already known ones, as well as with attempts by scientists to increase the objective component of the diagnosis, without the previous reliance on the experience gained by the practitioner. In addition, they are trying to create automated diagnostic systems that can work in the field and even in widespread and inexpensive devices, like smartwatches [2], which have in their arsenal a certain set of sensors for the physiological state of a person.

With the help of various kinds of portable devices, you can read (and, if necessary, write to the database files) an electroencephalogram (EEG), an electrocardiogram (ECG) or sound signals of the tones and noises of the heart – a phonocardiogram (PKG), body temperature, systolic and diastolic pressure, photoplethysmogram and other indicators When

organizing a signal processing procedure with a very small delay (practically "in real time"), it becomes possible to instantly respond to various deviations and malfunctions in the body and quickly stop them by changing the situation, taking medicines and using other modern medical methods and techniques [3].

It should be noted that to date there are a number of commercial software packages that allow processing data files according to various physiological indicators after the fact and making an alleged diagnosis (for example, the Easy ECG program is known among Russian products [4]), however, these programs are usually developed for a certain operating system, differ in low speed and are not suitable for use in mobile devices. Because of this, they are suitable only for diagnosing chronic diseases and for conducting scientific research by population groups, but are completely useless for a quick response to a change in the physiological state of a person.

The purpose of this work is to create a methodology for diagnosing deviations of physiological indicators of a person from his individual physiological norm according to PKG based on the development of automated software using modern algorithms.

Research methods and principles

Thanks to the authors' study of the state of the problem and some preparatory work, a research strategy was developed aimed at achieving the described goal, consisting of several stages:

1. Collection of a database with readings of sensors of various human health indicators (impersonal files, but with the obligatory indication of a real confirmed diagnosis).
2. Choosing the most appropriate programming language for research purposes.
3. Select a method and create a program for primary analysis of files from the database.
4. The use of neural network programming to identify the physiological state without the participation of specialists.
5. Approbation of the above steps on volunteers.

At the moment, some encouraging results have been obtained up to and including paragraph 3 of the strategy.

Many reference databases were found on the Internet, containing files with readings of various medical sensors, indicating the diagnoses detected in patients [5]. In addition, official negotiations are underway with the management of medical clinics in order to obtain additional databases (necessarily impersonal).

At the same time, the team of authors is intensively working on the creation of diagnostic software. This paper will describe the result achieved to date.

The multi-platform and freely distributed Python programming language, recognized as the most popular at the moment, is chosen as the language, which has high performance and almost unlimited possibilities for expanding its capabilities. In addition, for the Python language, there are libraries of special functions described below, which are necessary to implement the tasks set by the authors.

Let's consider the algorithm for Python developed by us, which performs processing of diagnostic files, consisting of the following blocks:

1. Import required function libraries: Scipy.io.wavfile to convert files downloaded from the wav database to data arrays [6]; Scipy.fftpack for frequency data analysis by FFT (fast Fourier analysis) [7]; Numpy for processing arrays of numerical data [8]; Matplotlib.pyplot for displaying graphical analysis results on the screen and in a file; Scipy.ndimage to create the maximum envelope of the resulting spectrogram (when analyzing the results of processing physiological data files).
2. Writing to the data arrays of the results of reading the loaded wav file of the oscillogram of the analyzed signal: digitization frequency and sample amplitudes (measurements at each digitization).
3. Filtering extraneous noise (zeroing samples with amplitudes of less than a certain value) and trimming the created data array to a given number of samples (for example, the sample files used in this study were cut to a duration of 10 seconds).
4. Calculate and display the sample time interval (reference information for process control).
5. Create data arrays for the spectrogram from the oscillogram data using the FFT (Fast Fourier Transform) method for subsequent frequency analysis.
6. Primary analysis of the obtained spectrogram by calculating the trend line (approximating the dependence of a given order) using the Numpy.polyfit function and outputting this line to the spectrogram graph.
7. Secondary analysis of the spectrogram using the Scipy.ndimage.maximum_filter function (maximum detection filter). In the future, it is planned to test other automated analysis methods.
8. Using the functions of the Matplotlib.pyplot module to display on the screen and in a new file simultaneously the results of processing the signal spectrogram read in paragraph 2, as well as its oscillogram with the results of primary and secondary analysis.

Main results

Let's consider an example and prospects for using the developed program code using the example of processing several files of the PKG database containing phonocardiograms of 1568 subjects in wav format, as well as a description of the diagnosis for each file in a text file supplied by a doctor. The dataset was collected as part of two mass screening campaigns conducted in Northeast Brazil in July-August 2014 and June-July 2015 [9]. Problems in the heart of subjects are systolic noise of different nature (four forms), called: crescendo, decrescendo, diamond and flat (plateau).

The mechanism of noise is explained by the appearance of turbulent blood flow in various areas of the heart muscle and the underlying vessels. They can occur in transient disorders of the blood flow, or in organic diseases of the heart. We have not yet considered this issue in more detail.

The task of demonstrating the developed method so far is only to identify the nature (form) of noise. To make a final medical diagnosis, additional large-scale studies will be carried out by a group of specialists, which includes the authors of this article.

From the described PKG database, 3 random files with different "shape" of noise were selected: decrescendo, plateau and diamond. All of them are measured in the same region of the human body – in the area of the tricuspid heart valve, located between the right ventricle and the right atrium. The function of the valve is to prevent reverse blood flow during the passage of systole. The noise in his work indicates an abnormal course of this process and may indicate a future, or an existing heart disease.

The program trims all PKG files to the same length of 10 s to ensure the correctness of the results obtained (initially the length of the signals was from 20 to 30 s or more).

Figure 1 shows the result of the program with recording the heartbeat with decrescendo noise.

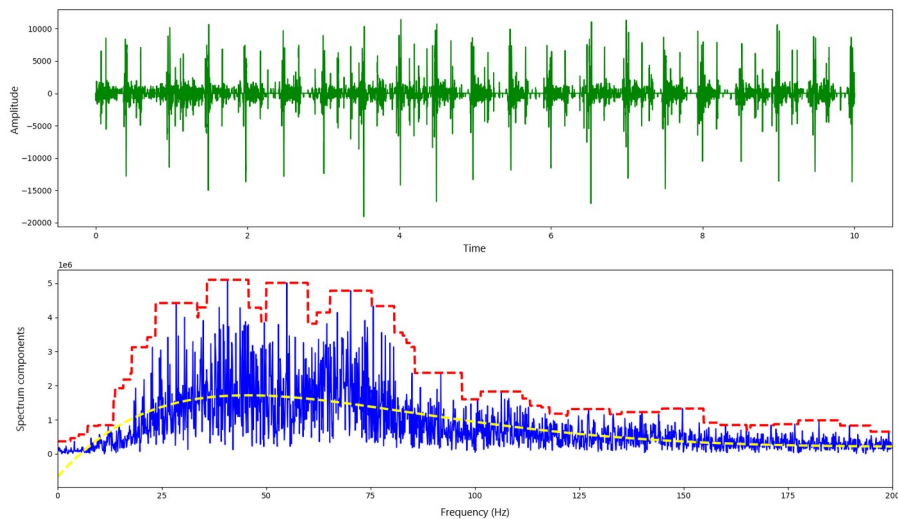


Figure 1 - Oscillogram, spectrogram and PKG processing results with decrescendo noise
DOI: <https://doi.org/10.23670/IRJ.2023.130.1.1>

The noise spectrograms presented in Figure 1 and all subsequent visualized ones are cut off (by setting the corresponding setting in the developed program) on the right at a frequency of 200 Hz, since preliminary analysis of the entire signal spectrum showed that above this frequency the components of the spectra of all the studied signals are almost zero. In addition, since the digitization frequency is 4000 Hz, according to Kotelnikov's theorem, only peaks explained by the imperfection of the FFT algorithm can be present at frequencies below 2000 Hz on the spectrogram.

Figure 1 clearly shows that the maximum spectrum amplitudes detected by the program are in the range of 25-75 Hz. Decrescendo noise, according to the theory, indicates that the patient has stenosis of the aortic mouth or pulmonary trunk.

Figure 2 shows the result of PKG processing with diamond noise.

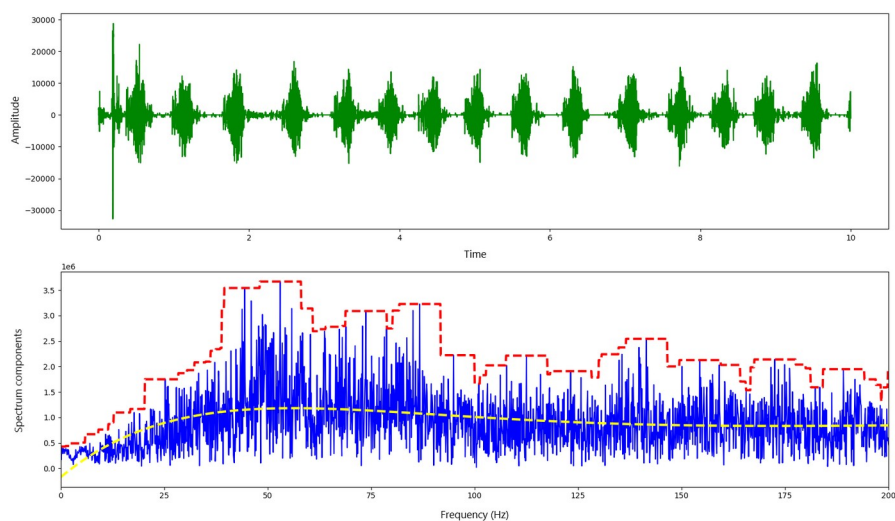


Figure 2 - Oscillogram, spectrogram and PKG processing results with diamond noise
DOI: <https://doi.org/10.23670/IRJ.2023.130.1.2>

Here we see a completely different picture of the analysis results – the noise frequency varies little in the frequency range of 20-200 Hz with the presence of some increase in amplitude at frequencies 40-60 and 70-90 Hz. But in general, in the entire

presented frequency range of 25-200 Hz, the noise amplitude does not change much, a maximum of 2 times. Such noise may indicate that the patient has anemia; its cause is a decrease in blood viscosity and, as a result, an increase in blood flow.

The result of processing the third PKG file is with noise of the "plateau" type, shown in Figure 3.

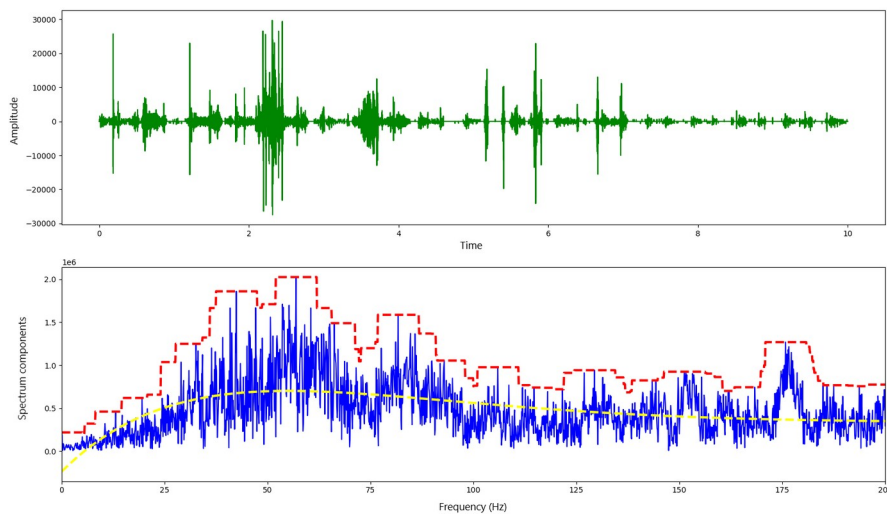


Figure 3 - Oscillogram, spectrogram and PKG processing results with plateau noise
DOI: <https://doi.org/10.23670/IRJ.2023.130.1.3>

In this case, in the range of 25-200 Hz, the amplitude of the frequency components of noise changes much more than in the case of the diamond noise study - up to 3 times. Also, at a frequency of 180 Hz, there is a pronounced peak of amplitude, and 3 narrow peaks are formed at frequencies of 35-45, 55-65 and 80-90 Hz. In general, the spectrogram of this type of noise from all those considered turned out to be the most complex, complex. This broadband type of heart noise is generally considered to be caused by transient functional reasons, for example, in adolescents.

Also, when analyzing all the presented records of cardiac noise, approximation dependencies of the spectrum of different orders were derived. Although they did not turn out to be useful in terms of identifying noise of different types, at least they made it possible to distinguish a spectrum range useful for diagnostics – up to 200 Hz. Interesting is the fact that the Numpy.polyfit function allows you to find the coefficients of the regression equation of very high orders - for example, in Figures 1-3, dashed yellow lines show graphs of 20th-order trend lines (with regression coefficients output to the program console), and this is not the limit. Also, in the case of insufficient order (maximum degree) of regression, i.e. with a large standard deviation, the program warns about this.

Discussion

We will determine the range of issues solved in this study, as well as the problems that are planned to be worked on.

The described study can be considered some test of the capabilities of the Python programming environment, during the application of which it was possible to solve all the questions arising in the course of the described work, therefore, it is quite perfect to achieve the goals of this study. It should also be noted the convenience of using libraries and special Python functions [10] - some results were obtained using only 1–2 commands. The created program code is executed very quickly and is quite suitable for processing signals "in real time."

Further, the maximum amplitudes of frequency components obtained during the production of spectrograms of processed signals and identified programmatically are planned to be used as new signals supplied to the neurons of the input layer of the neural network (the number of these neurons should be equal to the number of frequency components – you can limit, for example, to ten maxima). Then, experimentally, it is necessary to select the number of intermediate neural layers and the number of neurons in them. The output layer must contain a number of neurons equal to the number of diagnoses. The authors have experience with neural networks of different types and topologies.

At the next stage, the developed method should be tested on different sets of initial data: PKG, ECG, EEG and others, with a choice of sets that are most suitable for diagnosing a certain group of diseases or deviations of the physiological state of the body. It is also planned to use different types and models of sensors and primary transducers. Tests are required to be carried out separately in different age groups.

Also, with the help of the developed methodology, it is possible to identify not only diagnoses, but also the prerequisites for their occurrence. To do this, the measurement and processing of physiological signals should be carried out dynamically, continuously, preferably in real time. Another interesting task, possible for solving by the method under consideration, is the identification of the emotional background of people; at the same time, sensors that take physiological readings can be a thermometer, a tonometer, a photoplethysmograph and other sensors built into a smartwatch. It is also possible to create your own domestic model of such a watch, which contains a heart rate sensor in the design (a wide range of such portable sensors are available on sale in online stores).

Conclusion

The following results were obtained during the described work:

- the purpose is defined, and the study plan is drawn up;
- review of literature (including Internet sources on the topic of research);
- the programming language for writing program codes is selected;
- a test database with phonocardiograms of persons with different manifestations of systolic noise was selected;
- the algorithm of the program for PKG analysis was created, the functions required for this library were selected;
- program code is written that allows to implement the tasks planned by this study;
- with the help of the written program, a large number of FKGs from the database were analyzed, the results of the analysis of some phonocardiograms are presented in the article; briefly described are the prospective diagnoses of patients with a different spectrum of identified cardiac murmurs.

In addition, the prospects for the development of the method described in the article and its use in a wider range of innovative applications are described.

Конфликт интересов

Не указан.

Рецензия

Сообщество рецензентов Международного научно-исследовательского журнала
DOI: <https://doi.org/10.23670/IRJ.2023.130.1.4>

Conflict of Interest

None declared.

Review

International Research Journal Reviewers Community
DOI: <https://doi.org/10.23670/IRJ.2023.130.1.4>

Список литературы / References

1. Коваленко Н.М. Автоматизированные системы для диагностики уровня здоровья / Н.М. Коваленко, А.Ф. Неретина, А.Н. Пашков // Системный анализ и управление в биомедицинских системах. — 2010. — Т. 9. — 3. — с. 707-709.
2. He J. A New Approach for Daily Life Blood-pressure Estimation Using Smart Watch / J. He, L. Shu, X. Xu et al. // Biomedical Signal Processing and Control. — 2022. — Vol. 75. — p. 103616.
3. Gurina M.A. Assessment and Quality Management of Medical Center Services: Modern Approaches, Models and Methods / M.A. Gurina, Yu.V. Romyantseva, A.A. Wahi Nabil // Services in Russia and Abroad. — 2021. — Vol. 15. — 5(97). — p. 123-139.
4. Рябыкина Г.В. Дистанционная передача электрокардиограмм и системы централизованного анализа и архивирования электрокардиограмм. Опыт использования системы EASY ECG в ФГБУ РКНПК Минздрава России / Г.В. Рябыкина, Е.Ш. Кожемякина А.В. Соболев [и др.] // Терапевтический архив. — 2012. — Т. 84. — 4. — с. 52-57.
5. PhysioNet. Databases. — URL: <https://physionet.org/about/database> (accessed 28.11.2022).
6. Ильичев В.Ю. Использование библиотеки ZenCAD языка Python для разработки универсальной методики создания объемных изделий / В.Ю. Ильичев // Системный администратор. — 2021. — 6(223). — с. 82-85.
7. Amos Jeeva Oli H. Implementation of 4096 Point fft Using Modified 4 Point radix 2 FFT kernels to Reduce Latency / H. Amos Jeeva Oli // International Journal of Applied Engineering Research. — 2015. — Vol. 10. — 9. — p. 24147-24164.
8. Ilichev V.Yu. Development of Procedure for Determination of Characteristics of Heated Polycarbonate Greenhouses / V.Yu. Ilichev // International Research Journal. — 2021. — 2-1(104). — p. 132-135.
9. Oliveira J. The CirCor DigiScope Dataset: From Murmur Detection to Murmur Classification / J. Oliveira et al. // IEEE Journal of Biomedical and Health Informatics. — Vol. 26. — 6. — p. 2524-2535. — DOI: 10.1109/JBHI.2021.3137048.
10. Сафронова М.Е. Разработка методики использования библиотеки Graphviz для создания направленных графов / М.Е. Сафронова // E-Scio. — 2022. — 3(66). — с. 630-638.

Список литературы на английском языке / References in English

1. Kovalenko N.M. Avtomatizirovannyye sistemy dlja diagnostiki urovnja zdorov'ja [Automated Systems for Health Level Diagnostics] / N.M. Kovalenko, A.F. Neretina, A.N. Pashkov // Sistemnyj analiz i upravlenie v biomedicinskih sistemah [Systems Analysis and Management in Biomedical Systems]. — 2010. — Vol. 9. — 3. — p. 707-709. [in Russian]
2. He J. A New Approach for Daily Life Blood-pressure Estimation Using Smart Watch / J. He, L. Shu, X. Xu et al. // Biomedical Signal Processing and Control. — 2022. — Vol. 75. — p. 103616.
3. Gurina M.A. Assessment and Quality Management of Medical Center Services: Modern Approaches, Models and Methods / M.A. Gurina, Yu.V. Romyantseva, A.A. Wahi Nabil // Services in Russia and Abroad. — 2021. — Vol. 15. — 5(97). — p. 123-139.
4. Rjabykina G.V. Distancionnaja peredacha jelektrokardiogramm i sistemy centralizovannogo analiza i arhivirovanija jelektrokardiogramm. Opyt ispol'zovaniya sistemy EASY ECG v FGBU RKNPK Minzdravsocrazvitija Rossii [Remote Transmission of Electrocardiograms and Systems for Centralized Analysis and Archiving of Electrocardiograms. Experience of Using EASY ECG system in FGBU RKNPK of Ministry of Health and Social Development of Russia] / G.V. Rjabykina, E.Sh. Kozhemjakina A.V. Sobolev [et al.] // Terapevticheskij arhiv [Therapeutic Archive]. — 2012. — Vol. 84. — 4. — p. 52-57. [in Russian]
5. PhysioNet. Databases. — URL: <https://physionet.org/about/database> (accessed 28.11.2022).
6. Il'ichev V.Ju. Ispol'zovanie biblioteki ZenCAD jazyka Python dlja razrabotki universal'noj metodiki sozdaniya ob'emnyh izdelij [Using the ZenCAD Python Library to Develop a Universal Methodology for Creating Dimensional Products] / V.Ju. Il'ichev // Sistemnyj administrator [System Admin]. — 2021. — 6(223). — p. 82-85. [in Russian]

7. Amos Jeeva Oli H. Implementation of 4096 Point fft Using Modified 4 Point radix 2 FFT kernels to Reduce Latency / H. Amos Jeeva Oli // International Journal of Applied Engineering Research. — 2015. — Vol. 10. — 9. — p. 24147-24164.
8. Ilichev V.Yu. Development of Procedure for Determination of Characteristics of Heated Polycarbonate Greenhouses / V.Yu. Ilichev // International Research Journal. — 2021. — 2-1(104). — p. 132-135.
9. Oliveira J. The CirCor DigiScope Dataset: From Murmur Detection to Murmur Classification / J. Oliveira et al. // IEEE Journal of Biomedical and Health Informatics. — Vol. 26. — 6. — p. 2524-2535. — DOI: 10.1109/JBHI.2021.3137048.
10. Safronova M.E. Razrabotka metodiki ispol'zovanija biblioteki Graphviz dlja sozdaniya napravlennyh grafov [Developing a Methodology for Using the Graphviz Library to Create Directional Graphs] / M.E. Safronova // E-Scio. — 2022. — 3(66). — p. 630-638. [in Russian]