ANNOTATION
of the doctoral thesis for the degree of Doctor of Philosophy (PhD) in the educational program 8D06101 - «Information systems» (by industry)
Fedkin Yevgeniy

DEVELOPMENT OF MODELS, METHODS, AND ALGORITHMS FOR FORMING SCHEDULES FOR TRADITIONAL AND DISTANCE LEARNING IN THE UNIVERSITIES

The relevance of research. The task of synthesizing schedules has been the subject of scientific research since the mid-20th century. The scope of such tasks includes various fields of activity, including the preparation of lesson schedules. For most tasks, finding the optimal schedule is a difficult task, since these solutions must satisfy numerous constraints, many of which may have conflicts with each other.

Solving the problems of drawing up various types of schedules belongs to a special field of applied mathematics – the theory of schedules. This direction originates from the work of Henry Gantt, who proposed what is currently called "Gantt diagrams". The term "theory of schedules" was coined by R. Bellman in 1956.

One of the main issues of the "theory of schedules" is the classification and determination of the complexity of scheduling tasks. The most well-established classification was proposed by R. Graham. The currently available scheduling tasks can be classified into the following classes:

- NP-complete tasks;
- Polynomial tasks.
- For some types of tasks related to scheduling, it was not possible to determine the class of tasks to which they can be assigned.

To solve the problem of scheduling, various methods and algorithms based on various mathematical models are used. These mathematical models are based on the application of the target function and a set of constraints that are imposed on the final schedule. The mathematical models constructed in this way are further used for discrete optimization to build an optimal schedule.

The development of information technologies has led to their widespread use in education and the emergence of such a type of training as distance learning. The use of this type of training in universities can be used both for conducting certain types of classes together with full-time training and is the main form of training when training is conducted completely remotely. The expansion of the use of distance learning requires taking it into account in the educational process, including when drawing up lesson schedules.

The development of information management systems in universities poses the task of automating the scheduling of classes based on universal algorithms that can take into account the specifics of a particular educational institution for various forms of education.

Numerous studies have been devoted to the problem of scheduling both in the CIS (Klevansky N.N., Maslov M.G., Galuzin K.S., Nizamova G.F., Milekhina T.V., Asvad Firas M., etc.) and abroad (R. Lewis, M. Marte, H. Larget, M.T. Jensen, C. Mihaila,

etc.). Publications on this topic suggest using various methods and models: classical methods, metaheuristic, multi-agent systems, methods of solving by use cases. The market of automated systems is also in development and many well-known IT companies offer their developments in this area.

The relevance of the topic of this dissertation is determined by the need to create such methods, models and algorithms that can be used for scheduling in higher education institutions based on the specifics of the educational process of a particular educational institution. This dissertation work is devoted to the development of such methods, models and algorithms, their theoretical justification and practical application.

The object of the study is the process and its informational support, which takes place when drawing up schedules for various types of training (full-time, correspondence, distance, mixed, etc.) in higher educational institutions.

The subject of the study are models, methods and algorithms used in the process of scheduling for various types of education (full-time, correspondence, distance, mixed, etc.) in higher educational institutions.

The purpose of the research is to develop models, methods, and algorithms for scheduling in higher education institutions for various types of training (full-time, correspondence, distance, mixed, etc.) in higher education institutions, as well as to create an automated information system for scheduling classes based on the developed models, methods, and algorithms.

The objectives of this research include:

a) Analysis of existing approaches to scheduling

b) Development of models for the formation of class schedules

c) Development of methods and algorithms for scheduling based on the specified criteria

d) Development of an automated information system that includes

- A database for storing the data necessary for scheduling

- Software modules that implement database support and implement the developed models, methods, and algorithms for scheduling

e) Testing of the developed automated information system for scheduling classes.

Research methods. Various methods from schedule theory, graph theory, mathematical modeling, system analysis, computational complexity theory, dynamic programming method, as well as modern programming methods were used in this work.

The scientific novelty of the scientific study lies in the application of a modification of the genetic algorithm for scheduling classes using the adaptability function based on the implementation of restrictions using weight coefficients.

The practical significance of this work lies in the possibility of using the proposed algorithm for scheduling classes at universities, by speeding up the compilation and ensuring the optimality of the schedule.

The main scientific provisions submitted for defense:

a) the mathematical model of the task of scheduling classes based on the target function of evaluating the schedule based on the weighting coefficients of the restrictions imposed on the schedule and divided into "hard" and "soft" restrictions.

b) the algorithm for generating an initial population of schedules for a genetic algorithm based on clustering input data by academic flows and applying a target function to determine the optimal inclusion of an academic flow in the initial schedule.

c) algorithms of genetic operators (mutations and crosses) for the formation of a new population and selection of individuals for the formation of a new generation, as well as a condition for completing the search for a solution

d) the algorithm for constructing functions for evaluating the implementation of constraints with the possibility of optimizing calculations.

e) the architecture of the information system and the software implementation of the proposed methods and algorithms for scheduling classes.

Implementation of study results. The study was conducted on the basis of D.Serikbayev East Kazakhstan Technical University. Based on the data of the educational portal, the proposed algorithm for scheduling was tested. The proposed algorithm was used to schedule classes for the 3rd trimester of the 2022-2023 academic year. The proposed software implementation of scheduling has been implemented in the software package of the educational portal and will be used in the educational process.

Approbation of the study. The main results of the dissertation study were reported and discussed at the following international conferences:

- CITech-2018 - Computing and Information technologies in Science, technology and education, Ust-Kamenogorsk, Republic of Kazakhstan.

- SIBIRCON 2019 - International Multi-Conference on Engineering, Computer and Information Sciences, Novosibirsk, Russian Federation (indexed in the Scopus database).

- The 11th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications, IDAACS 2021, Krakow, Poland (indexed in the Scopus database).

The author's certificate No. 20431 dated September 23, 2022 "Program for scheduling semester classes for universities", authors Kumargazhanova S.K., Denisova N.F., Rakhmetullina S.Zh., Smailova S.S., Fedkin E.M., was received for the scheduling software package developed within the framework of scientific research.

Publications. 6 scientific publications have been published on the topic of scientific study:

a) 2 scientific articles indexed in the Scopus database:

- in the journal Acta Polytechnica Hungarica (75th percentile, 1st quartile)

- in the journal Eastern-European Journal of Enterprise Technologies (47th percentile, 3rd quartile).

b) a scientific article in the journal "Bulletin of the L.N. Gumilev Eurasian National University. Mathematics series. Computer science. Mechanics" is included in the list of recommended CCSON for the publication of scientific results.

c) monograph "Development of an on-line education model for higher educational institutions of the Republic of Kazakhstan", S.K. Kumargazhanova, S.S. Smailova, N.F. Denisova, S.Zh. Rakhmetullina, E.M. Fedkin, V.N. Zuev, L.Zh. Kakisheva, Ust-Kamenogorsk: VKTU, 2022. 146 p. ISBN 978-601-208-800-7.

d) 2 scientific articles in the scientific journals "Bulletin of the Academy of Civil Aviation" and the joint issue of the scientific journals "Bulletin of the D.Serikbayev East Kazakhstan State Technical University" and "Computing Technologies".

The structure and scope of the dissertation. The dissertation consists of an introduction, four chapters, a conclusion, a list of sources used. The total volume of the dissertation includes 234 pages.

The introduction substantiates the relevance of the research topic, the purpose, object, and subject of the study, defines the tasks and methods of research, scientific novelty, scientific provisions submitted for defense, the practical value of the research.

The first section of the dissertation provides an overview of the methods and algorithms used to solve the problem of scheduling classes. Conventionally described methods can be divided into 2 large groups: exact and approximate algorithm methods. Exact methods and algorithms assume a complete search of all possible combinations of solutions and the choice of the optimal one. The exact methods include branch and boundary methods, clipping methods, graph theory methods, etc. The applied methods are focused on obtaining a schedule close to optimal in an acceptable time based on a variety of heuristics. This group includes such methods as the ant colony method, the annealing simulation algorithm, constraint-based methods, neural networks, and genetic algorithms.

The second section describes the development of a mathematical model for scheduling classes. This model includes input data for the schedule: study groups and subgroups, teachers, classrooms, disciplines, types of classes, days, time intervals and study streams. The output data for the proposed model is a list of cards with the following set of values: training stream, audience, day, and time intervals. Restrictions are imposed on the compiled schedule, which are divided into "hard" (mandatory) and "soft" (recommendations). To find a solution to the problem, a modification of the genetic algorithm based on the adaptability function based on weighting coefficients and evaluation of the fulfillment of restrictions was used. As an individual representing the schedule, an individual with a single chromosome was selected, which has a set of genes representing individual classes. Each gene has an unchangeable part (academic flow) representing the description of the lesson in the schedule, and a changeable part representing how the described lesson should be included in the schedule (day, time, audience). The work of the proposed genetic algorithm is based on the generation of an initial population, considering the clustering of input training streams on their semantic similarity, then a cycle of applying genetic mutation and crossing operators, forming a new population, and checking the conditions for completing the search for a solution. Also in this section, a scheme for creating functions for evaluating constraints is proposed, which allows optimizing the number of calculations.

The third section of this paper provides a scheme for developing a database for storing source data for scheduling. This database includes tables for storing the source data described in the second section for scheduling. The physical implementation of the database is based on the Microsoft SQL Server 2017 database management system. Also in this section is a description of the software implementation of the dynamic library, which implements the genetic algorithm proposed in the second section for scheduling. This library includes support for input and output data for the schedule,

components for implementing constraint checking and saving the results in XML format. This library is extensible, which allows you to use it to implement scheduling for various educational institutions.

The fourth section describes the architecture of the educational portal of the D. Serikbayev EKTU and describes its components for the schedule of classes and the data used for scheduling. Based on the analysis carried out in this section, the implementation of data sampling from the database of the educational portal to the database for scheduling developed in the third section was carried out. To test the implemented dynamic library in the third section, based on the data of the educational portal, a schedule was generated for the 3rd trimester of the 2022-2023 academic year. A computational experiment was also performed to evaluate the proposed genetic algorithm. The specified computational experiment showed the following results: the proposed algorithm has a polynomial dependence on the number of elements in the final schedule, the proposed algorithm for generating the initial population allows you to get a solution close to optimal, which allows you to get a better final solution due to the cyclic part of the genetic algorithm.

In conclusion, a list of the theoretical and practical results of the study is provided, confirming the novelty and practical significance of the results obtained.

The dissertation ends with a list of sources used in writing the text of this dissertation.