IEEE Second International Conference on Data Stream Mining & Processing
August 21-25, 2018, Lviv, Ukraine

Educational Schedule Development Using Evolution Technologies

Oleh Suprun
Intellectual and Information Systems Department
Taras Shevchenko National University of Kyiv
Kyiv, Ukraine
oleh.o.suprun@gmail.com

Olena Sipko
Information Technologies Engineering Department
Cherkasy State Technological University
Cherkasy, Ukraine
barchat_08@mail.ru

Vitaliy Snytyuk
Intellectual and Information Systems Department
Taras Shevchenko National University of Kyiv
Kyiv, Ukraine
snytyuk@gmail.com

***Abstract* — An effective and optimal schedule is one of the keys to obtaining the necessary knowledge and skills by students, as well as creating comfortable conditions for teachers to work. However, in most universities, schedules are still being developed manually, which takes a long time, and always has a certain percentage of subjectivity. This is due to the relevant software inability to take into account the preferences and requirements of students and teachers and to give them the priorities as a human expert can do. The proposed method makes it possible, by conducting surveys among students and faculty members, to determine and take into account their requirements, to evaluate the possibility of fulfilling their desires and to prioritize them, depending on various factors such as material provision, the number of students with certain proposals for the schedule, teacher's position and status, and others. Using evolutionary technologies allows to quickly conduct an analysis that makes possible to perform multiply experiments, changing certain parameters, and choose the best option. The automation of this process guarantees taking into account all the restrictions and desires that a human expert can not handle with when dealing with a large number of students and teachers. This removes the influence of subjectivity. The proposed method was tested on real data, its efficiency and advantages are shown in the paper.**

Keywords—education process, schedule development, genetic algorithm, penalty functions.

I. INTRODUCTION

The schedule development problem is being solved by every person almost daily, regardless of whether it happens deliberately or not. Correct and optimal planning of future actions is one of the key factors affecting the final result of any process, it determines the efficiency and profitability o f the conducted economic and financial transactions.

Education process at universities is no exception. Creating a valid curriculum is an essential prerequisite for necessary skills acquisition by future specialists. This is especially important recently, due to the information society rapid development. As new in-depth disciplines and areas to study are emerging constantly, students have to learn more and more information. Also, such aspects as providing sufficient amount of time for learning, preparing for exams and rest, must be taken into account. The information understanding and overall training level depends on the timely submission of information, laboratory workshops organization, practical trainings and seminars. Correct schedule preparation is also equally important for teachers since it regulates the labour rhythm and directly affects their productivity and efficiency. Not to mention the universities technical provision, that is often far from ideal. The lack of audiences for the simultaneous placement of all students, computer equipment, or the necessary training material appears rather often.

Thus, the curriculum and schedule directly influences the level of specialist's final training, and some little at first glance errors or inaccuracies may have significant outcomes in the future.

Despite this, mathematicians began the necessary methodology development relatively recently. In 1967, in United States the world's first book on the theory of schedules was published [1]. One of the earliest conferences devoted exclusively to the problem of scheduling was the International Conference on Theory and Practice of Automated Timetables, held in Edinburgh in 1995 [2].

Nowadays, many literature on the curriculum development problem in higher education institutions can be found, but usually they are of a purely theoretical nature and are interesting only as a research of a non-trivial mathematical problem, or on the contrary - it has a narrow specialization, and the proposed method can be used only for a particular situation, such as the development of a curriculum for distance education [3], for software developers training [4,5], a schedule optimization for providing students with the necessary practical skills [6], or even to encourage the study of certain disciplines [7], and others.

Also many programs that allow to create schedule according to the specified rigid, that is technical, conditions, were developed. However, such programs do not allow to take into account the priorities and wishes of teachers and students, that is a negative psychological aspect. Most of the proposed programs often do not meet the requirements and are quite uncomfortable, therefore rarely used.

It is the subjectivism that is presented in the schedule creation process, which leads to numerous conflicts, significant time costs and appearance of suboptimal by different criteria decisions. The automation of the scheduling problem process is rather complicated problem, its algorithmization encounters aspects of NP (nondeterministically polynomial) complication. The search for precise algorithms for solving these problems, the time of which is limited by the polynomial of the input data size, can't give any proper results. Exhibitory selection algorithms require significant computational cost even when solving average dimensional examples. Therefore, one of the important areas of research is the construction and analysis of approximate algorithms with guaranteed accuracy estimation for NP-complete problems.

II. Problem Analysis

Solving the problem of schedule development in general is the process of executing some fixed tasks system using a certain set of resources or service devices. When transferring the general schedules theory to the educational and training schedule, the formulation of this problem class is as follows: "For a given set of training classrooms and a given set of time intervals (lessons), build such a distribution of training sessions for all objects (teachers and training groups) for which the chosen criterion of optimality is the best".

The educational schedule must fulfil the following basic requirements:

- precise and full schedule compliance with the curriculum by volume, content, type and time of classes, provision of materials for the curriculum and programs, systematic and continuous learning process throughout the day, and the evenly distributed students' work during the week, month and semester;
- providing on the classes of interdisciplinary and internal logical connections for each discipline, which are determined by its structural-logical scheme;
- providing of necessary time intervals for students to work independently between lectures and practical classes for each discipline, alternating between disciplines with different complexity levels and classes types during the day and the week;
- implementation of principles for teacher's and student's scientific work organization, achievement of an equal teachers employment in order to ensure their preparation before classes, systematic conducting of methodological, publishing and research activity (here the individual needs department head recommendations may be taken into account);
- providing the effective use of the auditor's and the training laboratory base.

Also there are other requirements of ergonomic and organizational-methodical nature. For example, they include the reasonable execution of the individual teachers preferences, the implementation of the individual training principle, conducting classes with small students groups, and others.

Developing a schedule, the problem of optimal resource management, such as the teaching staff and the auditorium, appears. Solving this problem, it is necessary to take into account the strict restrictions, as well as additional requirements that may be violated in some cases.

Strict restrictions are limitations that must surely be fulfilled; those that physically can not be violated. As example of these: "At the same time, the same audience should have only one teacher and one subject." As a result of solving this problem, it is necessary to get a schedule that simultaneously satisfies all the strict restrictions. If this is not possible, then the list of such restrictions should be changed or some measures should be taken to allow for an acceptable schedule.

Weak restrictions are limitations that can be violated, but this violation must be kept to a minimum level. Their performance is not as obligatory as the strict ones. In contrast to the strict restrictions that have an objective nature, soft restrictions are subjective. Thus, the restriction "The lesson must not be conducted in the laboratory" is objective-subjective, and the restriction "Lecturer Goroshko's lectures should be conducted on Monday and Tuesday" - subjective. It is obvious that violation of weak restrictions leads to the schedule deterioration, but does not exclude its admissibility. Since such violations can be many and they are of a versatile nature, the relevance of obtaining an acceptable optimal (acceptable) timetable is indisputable.

III. Problem formalization and objective function

Let the set $R=\{r_1, r_2, ..., r_n\}$ be a finite set of all possible schedules for a certain educational institution. Its finiteness is guaranteed by the finiteness of educational disciplines set $P=\{p_1, p_2, ..., p_m\}$, teachers set finiteness $L=\{l_1, l_2, ..., l_k\}$, students set finiteness $S=\{s_1, s_2, ..., s_k\}$ and auditoria's set $A=\{a_1, a_2, ..., a_v\}$.

The schedule development problem, without generality limitation, can be presented as follows [9]:

$$\max_{r\in R} F(r), r\in \Omega(P,S,L,A)\in R,$$

where Ω – a set of restrictions that are determined by the auditoria's and classrooms presence and specialization, the teachers distribution according to the disciplines, discipline according to the classrooms, etc.

Taking into account the requirements and priorities of teachers and students separately, the following formula can be obtained:

$$\alpha_S F_S(r) + \alpha_L F_L(r) \to \max, r\in \Omega(P,S,L,A),$$

where F_S – students objective function, F_L – teachers objective function, α_S i α_L are weighting coefficients indicating the priorities of teachers and students as the educational process subjects.

Considering the student as the dominant subject in a higher educational institution, it is rational to establish a priori $\alpha_S = 0,6$, $\alpha_L = 0,4$.

For further correction, the following rule is used: if the ratio of the students' number to the teachers' number corresponds to the normative value, the values of the coefficients do not change, if the real ratio is different from the normative, then α_S and α_L must be corrected.

Let N_S be the students number, N_L – teachers number, Nom – the nominal value of students number ratio to the teachers number determined by the managing authority. If inequality $\frac{1}{2}Nom \le \frac{N_S}{N_L} \le Nom,$ is true, then

$$\alpha_S = 0,6 - \frac{1}{2} Nom\left(\frac{N_S}{N_L} - \frac{1}{2} Nom\right) \cdot 0,4 .$$

The features of teachers' and students' preferences and requirements set forming are presented below.

Obviously, students can be considered as a certain set, divided into classes (groups according to courses and specialties). Students in each group independently formulate their schedule requirements, which form a single list. If it meets similar requirements - they are united. The authors of the opposing requirements are offered to reach the agreement or to withdraw their claims altogether. In the event of their disagreement among the students, the vote is conducted and the requirement is chosen by the majority.

Thus, a set of requirements is obtained Z^v , it has different priority for each student. In order to reconcile individual preferences, each student is given the opportunity to determine the advantage of each requirement. For this purpose, the hierarchies analysis method is used [8], the matrices of pairwise comparisons are constructed, for which the maximum eigenvalues and corresponding vectors are found. Let the $\lambda_{\max}$ be the eigenvalues, for the i -th student, $i = \overline{1,m}$, $x^i = \left(x_1^i, x_2^i, ..., x_l^i\right)$ – corresponding eigenvector. Performing the normalization of this vector elements by the formula

$$x_j^{iн} = \frac{x_j^i}{\sum_{j=1}^{l} x_j^i}$$

is obtained, that $x_j^{iн} \in (0,1)$ and $\sum_{j=1}^{l} x_j^{iн} = 1$. It can be said that the value $x_j^{iн}$ shows a priority of j -th criteria for i -th student. Since all students are equivalent (equally competent), the requirements priority for them (most often, for the group) is defined as the average value of the criteria priorities for each student, that is:

$$x_j = \frac{1}{m} \sum_{i=1}^{m} x_j^{iн}, \; j = \overline{1,l} \; .$$

Thus, for the students group, a requirements priority vector is obtained: $X = (x_1, x_2, ..., x_l)$. Objective formula can be modified:

$$\alpha_s \cdot F_s = \alpha_s \cdot \sum_{j=1}^{l} x_j \cdot \chi\left\{Z_j^v\right\},$$

$$\chi\{A\} = \begin{cases} 1, \text{if } A \text{ is true}, \\ 0, \text{otherwise}. \end{cases}$$

Teachers do not have groups and their requirements need to be individually implemented. Let M be the number of teachers, divided into sets $T = \{T_1, T_2, ..., T_K\}$ (department head, doctors of sciences, assistant professors, assistants, etc.).

The specialist, who makes the schedule, forms a matrix of pairwise comparisons: defines the priorities of teachers, representatives of groups:

$$y = \{y_1, y_2, ..., y_K\}, \qquad y_i \in (0,1), \qquad \sum_{i=1}^{K} y_i = 1,$$

Each teacher has his own advantages in forming a schedule, and the number of such advantages from different teachers will be different. Let $Z_i^T = \left\{Z_{i_1}^{T_j}, Z_{i_2}^{T_j}, ..., Z_{i_{n_i}}^{T_j}\right\}$ be the advantages vector for i -th teacher from j -th set, n_i – number if its elements, $j = \overline{1,M}$, $i = \overline{1,K}$. Vectors Z_i^T will correspond the values of the priority vector calculated using the method given above $D_i^j = \left\{d_{i_1}^j, d_{i_2}^j, ..., d_{i_{n_i}}^j\right\}$, $i = \overline{1,M}$, $j = \overline{1,K}$. Thus:

$$\alpha_L F_L = \sum_{j=1}^{K} y_j \cdot \sum_{i=1}^{M} \chi\left\{L_i \in T_j\right\} \cdot \sum_{l=1}^{n_i} d_{il}^j \cdot \chi\left\{Z_{il}^{T_j}\right\},$$

$$\chi\left\{L_i \in T_j\right\} = \begin{cases} 1, \text{if } L_i \text{ belongs to } T_j, \\ 0, \text{ otherwise}. \end{cases}$$

The objective function can be rewritten as follows:

$$F(r) = \alpha_S \sum_{j=1}^{l} x_j \chi\left\{Z_j^v\right\} +$$

$$+\alpha_L \sum_{j=1}^{K} y_j \sum_{i=1}^{M} \chi\left\{L_i \in T_j\right\} \sum_{l=1}^{n_i} d_{il}^j \cdot \chi\left\{Z_{il}^{T_j}\right\} \rightarrow \max,$$

$$r \in \Omega(P, S, L, A).$$

where x_j and y_j are students and teachers preferences' priorities, Z_j^v students preferences, L_i teachers, T_j teachers' groups, $Z_{il}^{T_j}$ teachers' advantages, d_{il}^j - priority of these advantages, l number of students' preferences, K number of teachers' groups, determined by their positions, scientific degrees and academic rank, M number of teachers' groups, n_i number of teachers' in i -th group.

IV. Matrix-evolutionary method

In order for better visualization, the schedule representation can be shown as a rectangular parallelepiped (Fig. 1).

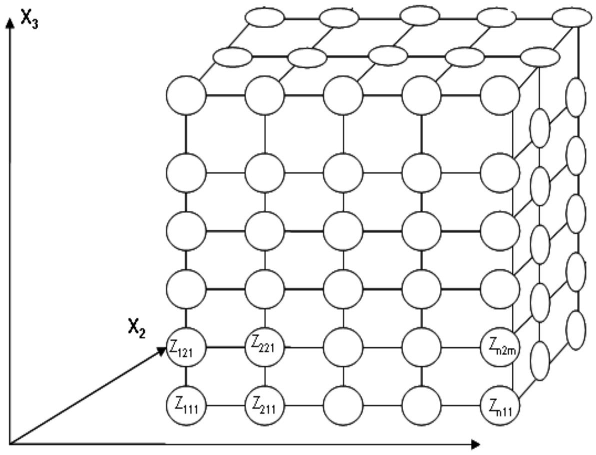

Fig. 1. Schedule representation.

where X_1 is day-lesson, X_2 - course-group, X_3 - auditory, Z - teacher-subject

Since the problem of automated schedule development is an NP-complete task, it is expedient to use the evolutionary technologies algorithm, namely, a modified genetic algorithm. In this case, the method of penalty functions is used, which makes it possible to transform a problem with constraints into a sequence of unconditional optimization problems of some additional functions [10]. They are obtained by modifying the target function with the help of restriction functions in such a way that limitations are not presented in the optimization problem in explicit form.

The modified method of solving the schedule development problem, using the penalty function, has the following steps:

Step 1. Define structure S of potential schedule r.
Step 2. Define the criteria E for the search stop.
Step 3. Perform a potential solution encoding.
Step 4. Until criteria E is done:
Step 4.1. Until a sample of potential solutions Z incomplete:
Step 4.1.1. Generate potential solution r.
Step 4.1.2. If it is unacceptable ($r \notin \Omega_1(P,S,L,A)$ go to Step 4.1.1.
Step 4.1.3. If it is acceptable ($r \in \Omega_2(P,S,L,A)$), add it to Z and go to Step 4.1.
Step 4.1.4. If the solution r is unacceptable according to at least one of $\Omega_2(P,S,L,A)$, one of three variants is done:
A: If to Step 4.1.1.
B: If variant A was made more than $A_{\max}$ times, go to variant C.
C: Set

$$F(r) = \alpha_S F_1(r) + \alpha_L F_2(r) - \beta_{it}\phi(F_1(r) \vee F_2(r)),$$

where β – weight coefficient, it – iteration number, $\phi(*)$ – penalty function. Consider r as potential solution and go to Step 4.1.1.

Step 4.2. For all potential solutions calculate $F(*)$, taking into account, that if the solution is acceptable, $\phi(*)$.

Step 4.3. Generate new potential solutions based on the values of the target function, using crossover operations and mutations (if the optimization method is a genetic algorithm) or using normally distributed numbers if this is an evolutionary strategy.

Step 5. Calculate the criteria E.

Penalty function construction is presented below.

$$\phi(F_1(r) \vee F_2(r)) = \begin{cases} 1, \text{ if } r \in \Omega_2(P,R,L,A) \\ f(\mathop{\vee}_{j=1}^{l} x_j \mathop{\vee}_{j=1}^{K} y_j .D, \gamma), \text{ otherwise} \end{cases}.$$

The zero value of a penalty function presents the situation where at least one student's requirement or at least one teacher's requirement is not fulfilled, D – value that integrates the teachers' requirements priorities, γ – penalty parameter.

Since "fined" may be solutions that do not fulfil the requirements of student groups, and solutions that do not fulfil the individual teachers' requirements, it is expedient to consider the penalty function additive and to write in this form:

$$\phi(F_1(r) \vee F_2(r)) = \beta_1\phi_1(F_1(r)) + \beta_2\phi_2(F_2(r)) =$$

$$= \beta_1 f_1(\mathop{\vee}_{j=1}^{l} x_j) + \beta_2 f_2(\mathop{\vee}_{j=1}^{K} y_j, D).$$

Obviously, the more restrictions are violated, the greater the penalty function value is. According to the construction, penalty function is an integral function, and based on its purpose, the inequality is valid:

$$0 \le \phi(*) \le F_{\max}.$$

Similarly, for its component: $0 \le \phi_i(F_i(r)) \le F_{i\max}$, $0 \le f_1(\mathop{\vee}_{j=1}^{l} x_j) \le F_{1\max}$ and $0 \le f_2(\mathop{\vee}_{j=1}^{K} y_j) \le F_{2\max}$.

V. Experimental results

To check the proposed algorithm, an automated system was created that develops schedule in accordance with the requirements of the educational process, the disciplines sequence correctness, and the desired of teachers and students. The system has been tested both on specially created and real data.

The system kernel and the interface part were written in the programming language Delphi 7.0. The presented solution is performed using object-oriented technologies, which will allow to easily encapsulate them in future in new system modifications, without violating the algorithms integrity. The database was implemented on the InterBase 6.0.

As real testing data were used information about students groups, teachers and subjects of full-time studying at the Faculty of Information Technologies of the Cherkasy State Technological University, as well as randomly generated initial data (disciplines were randomly assigned to the classes).

To determine the developed models and target functions effectiveness and expediency, as well as the created automated system relevance, calculations were made on the schedules created automatically and manually. The research was conducted on schedules for the Faculty of Information Technologies and Systems of Cherkasy State Technological University, created on various samples of the initial data: for different semesters (autumn, spring), for different academic years (2014-2015, 2015-2016) The target function results for different initial data are shown in Table I.

TABLE I. OBJECTIVE FUNCTION VELUES

Creation method	Autumn semester 2014-2015	Spring semester 2014-2015	Autumn semester 2015-2016	Spring semester 2015-2016
Manually	643	488	604	**543**
Automated	682	521	663	**571**
Difference in percents	**≈5,7%**	**≈6,3%**	**≈8,9%**	**≈4,9%**

The calculation of the target functions values obtained with automatic and manually creating a schedule methods showed that the effectiveness of the developed models and methods is about 5-9% comparing to the manually schedule creation for classes in higher educational institutions.

In addition to testing the automatic system optimality, a test was made on the dependence of the time, used for creating the optimal schedule, to the problem dimensionality. On average, from 5 to 10 tests for each dimensionality of the initial data was made. The result data are presented in the Table II.

TABLE II. CALCULATION TIME DEPENDING ON PROBLEM DIMENSIONALITY

Problem dimensionality (lessons number)*(groups number)*(days of week number)	Calculation time		
	max	min	**average**
5	0,8	0,05	**0,24**
25	3,2	0,9	**1,96**
45	5,4	2,1	**3,5**
65	12	3,1	**5,8**
85	14	5,2	**7,6**
105	25	10	**14,05**
125	39	14,5	**18,1**
145	46	19	**26,5**
165	**51**	**25**	**32**

In Figure 2 the dependence of the time used for solving the problem on the problem dimensionality is shown (the number of pairs per week and the number of groups).

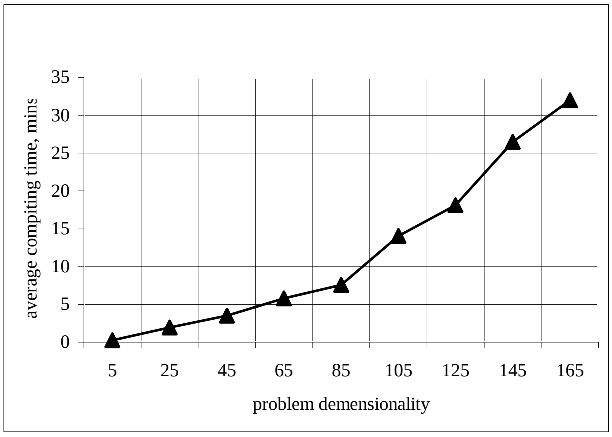

Fig 2. Calculation time.

As can be seen, the problem solving time increases with the increase of input data amount. This is due to a fast increase in the number of restrictions in the model, which increases the size of the arrays and, accordingly, the time used to solve the problem.

VI. CONCLUSIONS

Despite the existence of many programs for automated schedule development in higher education institutions, in most cases the schedule is formed manually, which requires considerable effort, it is time consuming, and such a schedule is far from always optimal and effective. This is due to the inability of such software to take into account the large number of preferences of students and teachers, to choose acceptable and unfulfilled preferences, and to take into account their priority.

The proposed schedule development problem formalization, and the algorithm for its direct solving, allows using the hierarchies' analysis method, conducting surveys among students and teachers, formalizing their preferences, and appropriately taking them into account, when developing a schedule. On the one hand, it can not completely replace the expert's analytical ability, based on his own experience and knowledge of the task, but at the same time, it will avoid subjectivity. In addition, an automated schedule creating system can take into account absolutely all preferences, even if some will be inappropriate, and optimise the schedule, while the specialist will not be able to effectively evaluate all the requirements and wishes, which is especially important for institutions with a large number of students.

The search for the optimal schedule is performed using a modified genetic algorithm, which has demonstrated itself well solving the optimization problems, which can not be solved by classical methods. It allows to find the optimal, or at least acceptable, solution using a small period of time.

Taking into account the preferences of students and teachers is made using the penalty functions, which allows to take into account the priority of these preferences, to evaluate their importance in general, and in comparison with each other. Due to this, the preferences and requirements are perceived more effectively than when manually developing the schedule, the optimal schedule contains the least amount of violated desires. Also, using the penalty method allows to simplify the target function, which generally accelerates the algorithm.

Experimental researches were done using the developed system for automated schedule creating, while both theoretical and real data were used. A comparison of the obtained schedule and the time used for its creation, with the schedules obtained by classical methods, in particular, generated manually by an expert, is performed, and the results are presented.

REFERENCES

[1] E. Balas “A Note on the Branch-and-Bound Principle,” Operations Research. New York, vol. 16, iss. 2, pp. 442-445, April 1968.

[2] E. Burk, and P. Ross, Practice and Theory of Automated Timetabling: First International Conference. Edinburgh: UK, 1995.

[3] Y. Yang, and D. Tian, “The research of multimedia curriculum design and curriculum development in modern distance education”, *2016 International Conference on Educational Innovation through Technology (EITT),* September 2016. Published in: 2017 3rd IEEE International Conference on Computer and Communications (ICCC), 13-16 Dec. 2017. DOI: 10.1109/CompComm.2017.8322799

[4] E. S. Grant, “A proposal for technology-based software engineering curriculum development,” *2016 IEEE 8th International Conference on Engineering Education (ICEED)*, Kuala Lumpur, Malaysia, December 2016.

[5] E. S. Grant and V. Shankararaman, “Technology-driven software engineering curriculum development,” *2014 IEEE 27th Conference on Software Engineering Education and Training (CSEE&T)*, Klagenfurt, Austria, pp. 168-170, April 2014.

[6] H. Fugang, C. Le, and Z. Qinhua. “Curriculum Development for Practical Training: A Design-Based Research of Network Detection Combat,” *2016 International Conference on Educational Innovation through Technology (EITT)*, Tainan, Taiwan, September 2016.

[7] W. Yoo, S. Pattaparla, and S. Shaik, “Curriculum development for computing education academy to enhance high school students' interest in computing”, *Integrated STEM Education Conference (ISEC), 2016 IEEE*, Princeton, NJ, USA, pp. 282 - 284 April 2016.

[8] T. L. Saaty, and Kirti Peniwati, Group Decision Making: Drawing out and Reconciling Differences. Pittsburgh, Pennsylvania, 2008.

[9] O. Sipko, and V. Y. Snytyuk, “Aspects of formulation of the objective function in the problem of scheduling in higher educational institutions based on subjective preferences,” Nauka i Studia. Polska, Przemyśl: Sp. zoo «Nauka i studia», no. 15 (125), pp. 39-51, 2014.

[10] V. Y. Snytyuk, and O. M. Sipko “Penalty Functions Using in the Schedule Development Problem,” Mathematical Machines and Systems, Kyiv, no. 3, pp. 158-164, 2015.

.