

УДК 681.3.06

О.М. Бойко

Черкаський державний технологічний університет, м. Черкаси, Україна
boyko-e-n@mail.ru

Еволюційна технологія розв'язування задачі складання розкладів навчальних занять

У статті викладено обґрунтування необхідності створення автоматизованої системи складання розкладу занять і аналітичний огляд класичних методів розв'язання задачі складання розкладів. Як предметна область розглянутий вищий навчальний заклад і проблеми складання розкладу навчальних занять студентів. Велика увага приділена використанню еволюційних методів для розв'язання задачі складання розкладів і дослідженню результатів роботи створеної системи.

Вступ

Кінцевим етапом планування навчального процесу у вищому навчальному закладі (ВНЗ) є складання розкладу занять. Розклад буде оптимально складеним, якщо він повністю реалізує взаємозв'язок навчальних дисциплін під час їх вивчення, задовольняє певним обмеженням та враховує педагогічний і методичний досвід викладачів. Якщо при складанні розкладу занять не враховані в повному обсязі зв'язки між предметами або не задоволені обмеження, то якість планування навчального процесу знижується, що істотно позначається на якості підготовки фахівців.

Розклад навчальних занять – це документ, що регламентує трудовий ритм, впливає на творчу віддачу викладачів, тому його можна розглядати як фактор оптимізації навчального процесу. Розклад повинен задовольняти педагогічним вимогам, що базуються на принципах аналітичної дидактики і кібернетичної аналогії [1].

Задачу складання розкладу не варто розглядати лише як деяку програму, що реалізує функцію механічного розподілу занять на початку семестру, на чому її (програми) використання й закінчується. Економічний ефект від ефективного використання трудових ресурсів може бути досягнутий лише у результаті кропіткої роботи з їх управління. Розклад є лише інструментом такого управління, і для найбільш повного його використання необхідно, щоб програмно-алгоритмічне забезпечення для складання розкладу об'єднало у собі засоби для підтримки його оптимальності у випадку зміни деяких вихідних даних, що на момент складання розкладу вважалися постійними. Крім цього, оптимальне управління такою складною системою неможливе без нагромадження статистичної інформації про процеси, що відбуваються у системі. Тому саме розв'язання задачі складання оптимального розкладу є лише частиною складної системи управління навчальним процесом.

Постановка задачі. Розв'язання задачі складання розкладів у загальній постановці є процесом виконання деякої фіксованої системи завдань за допомогою певної множини ресурсів чи обслуговуючих пристроїв [2]. При перенесенні загальної теорії розкладів на розклад навчальних занять формулювання задачі складання розкладу занять звучить так: «Для заданого набору навчальних аудиторій

(в даному випадку під навчальною аудиторією розуміється широке коло приміщень, в яких проводяться заняття (від комп'ютерної аудиторії до спортивного залу)) і заданого набору часових інтервалів (уроків або навчальних пар) побудувати такий розподіл навчальних занять для всіх об'єктів (вчителів і навчальних груп), для якого обраний критерій оптимальності є якнайкращим».

У статті запропоновано створення такої математичної моделі розкладу занять у ВНЗ, яка дозволяла б ефективно (у задані терміни і з заданою мірою оптимальності) розв'язувати задачу автоматичного складання розкладу занять і була б гнучкою (властивість незначних структурних змін вихідної інформації) у випадку адаптації системи для розв'язання конкретної практичної задачі.

Для побудови математичної моделі розкладу занять у ВНЗ введемо змінні і визначимо обмеження. Нехай у ВНЗ є G навчальних груп, k – номер навчальної групи ($k = 1, \dots, G$). Заняття проводяться в робочі дні в півторагодинні інтервали, які називають парами. Позначимо t – номер робочого дня тижня ($t \in T$), де T – множина номерів робочих днів для групи k ; j – номер пари ($j = 1, \dots, J$); J – загальна кількість навчальних пар. З кожною навчальною групою k протягом тижня відповідно до навчального плану проводиться S занять. Позначимо s – номер дисципліни в списку занять для групи k ($s = 1, \dots, S$). Передбачається, що заняття проводяться у всіх групах. Тоді, якщо з дисципліни протягом тижня проводиться більше одного заняття, ця дисципліна згадується в списку занять стільки разів, скільки їх передбачається навчальним планом для кожної групи. Нехай p – номер викладача ($p = 1, \dots, P$). Заняття кожного потоку можуть проводитися в наявному аудиторному фонді ВНЗ, який визначається до початку складання розкладу. Для того щоб повноцінно вести наукову, учбово-методичну роботу, готуватися до занять, викладач ВНЗ повинен мати вільний час. Ця умова не є достатньою, але є необхідною. Введемо вагові коефіцієнти, за допомогою яких повинен враховуватися статус викладача – його вчений ступінь і звання, займана посада, науково-суспільна активність і т.п. Пріоритетне право викладачів на розподіл занять в алгоритмі складання розкладу визначається за методом ранжування посад [3].

Оптимально складений розклад навчальних занять не повинен змінюватися під час семестру чи навчального циклу, щоб не порушити враховані в розкладі обмеження та зв'язки між об'єктами. Планування послідовності проведення видів занять з дисципліни в навчальному процесі необхідно розглядати з боку студента і його психофізіологічних особливостей і розміщувати заняття так, щоб практичні заняття проводилися після лекцій, а лабораторні заняття – після практичних. Це підвищить ефективність використання навчального часу, що відводиться державним освітнім стандартом вищої професійної освіти на вивчення окремих дисциплін.

Основною задачею при складанні розкладу є планування і забезпечення методично правильного процесу вивчення усіх навчальних дисциплін навчального плану: їх взаємозв'язки, правильна послідовність і чергування усіх форм навчальної роботи з дисциплін на базі врахування можливостей студентів зі сприйняття і переробки навчальної інформації.

Аспекти і проблеми застосування еволюційного моделювання

Існує багато класичних методів розв'язання задачі складання розкладу. Так, модель задачі складання розкладу в рамках лінійного цілочисельного програмування [4]

містить ряд недоліків, пов'язаних як з неповною адекватністю представленого розв'язку задачі, так і значною трудомісткістю використання запропонованого комплексу програм, що вимагає участі кваліфікованого користувача. Метод імітації випалювання [5] та алгоритм розфарбовування графу [6], незважаючи на зовнішню простоту, можуть виявитися цілком ефективними для складання лише невеликих розкладів. При реалізації алгоритму, що базується на принципах імітаційного моделювання [7], [8], обмежується можливість застосування розробленої системи в інших ВНЗ, крім того, знадобиться вносити істотні зміни в алгоритм при незначних внутрішніх змінах у ВНЗ. Усі ці методи в своїй основі використовують ітераційну техніку неперервної оптимізації. Очевидно, що вони орієнтовані на пошук лише локальних оптимумів, а глобальний оптимум може бути знайдений лише випадково. У зв'язку з цим раціонально використовувати методи, які зберігають переваги класичних і вільні від їх недоліків. До таких методів відносять еволюційне моделювання.

Загальний алгоритм складання розкладу занять за допомогою еволюційного моделювання складається з декількох етапів. Першим кроком при розробці математичної моделі є розробка структури хромосоми, в якій буде зберігатися розв'язок. У нашому випадку такою «хромосомою» є розклад. Обрана структура повинна враховувати всі особливості й обмеження шуканого розв'язку. На наступному кроці з хромосом формують початкову популяцію. Для створення нового покоління з кращими показниками із сукупності хромосом необхідно вибрати кращі для подальшої репродукції. Відбір хромосом здійснюється за допомогою цільової функції, яка визначає ступінь придатності хромосоми до формування наступного покоління. Щоб отримати нову популяцію, відібрані хромосоми попарно схрещують між собою, після чого до популяції застосовують оператор мутації. Після реалізації мутації цільова функція перераховується і все починається спочатку.

Таким чином, помістивши початкову популяцію у створене штучне середовище і реалізувавши процеси селекції, кросоверу і мутації, ми одержимо ітераційний алгоритм пошуку оптимального розв'язку. Розглянутий алгоритм не лише стійкий до локальних мінімумів, але й забезпечує відносно швидкий пошук оптимального розв'язку.

Але застосування еволюційного методу для розв'язання задачі складання розкладу викликає ряд проблем:

1. Формування хромосоми.
2. Визначення цільових функцій та обмежень.

Формування структури «хромосоми»-розв'язку задачі

Хромосома є набором генів (фрагментів розкладу (ФР)). Хромосоми представляють динамічним масивом, кожним елементом якого є ФР. Один фрагмент в хромосомі – це одне заняття в розкладі. Наш ФР матиме структуру, що складається з чотирьох частин закодованої інформації:

- перша частина – день тижня;
- друга – номер навчальної пари;
- третя – номер групи, викладача і предмету;
- четверта – номер аудиторії.

Розглянемо приклад, де початковими даними для складання розкладу занять є: 5 днів тижня; 5 пар на день; 20 навчальних груп; 40 викладачів; 80 предметів; 100 аудиторій.

ФР, зазвичай, є бітовим рядком фіксованої довжини, на кодування якого (в нашому випадку) відводиться 32 біти, а саме: на кодування першої частини фрагмента відводиться 3 біти (5 днів тижня), другої – 3 біти (5 пар в день), третьої – 18 біт (20 груп, 40 викладачів, 80 предметів), четвертої – 7 бітів (100 аудиторій). І таких фрагментів буде 500 (5 днів по 5 пар у 20 групах), тобто хромосома матиме 16000 біт. Наприклад, один фрагмент матиме вигляд: 010 011 10010 011000 0101110 1100001.

Якщо проаналізувати описане кодування, то можна побачити ряд недоліків, які, в свою чергу, істотно впливають на швидкість, а іноді й на збіжність алгоритму. По-перше, при такому кодуванні сусідні числа можуть відрізнятися в значеннях декількох бітів, так, наприклад, числа 7 (0111) і 8 (1000) у бітовому представленні розрізняються в 4-х позиціях, що ускладнює функціонування генетичного алгоритму і збільшує час його збіжності. Для того щоб уникнути цієї проблеми, можна використовувати кодування, при якому сусідні числа відрізняються меншою кількістю позицій, наприклад, використовуючи код Грея [9], але він також не є універсальним. По-друге, кодування такого роду призводить до невиправданої інформаційної надлишковості. На прикладі це можна пояснити так: використовуючи вищезгадані вихідні дані, для кодування першої частини ФР (дня тижня) нам знадобиться 3 біти, якими можна закодувати 7 чисел, але даних у нас менше – 5. Така надлишковість позначається на швидкості збіжності алгоритму, адже постійно потрібно перевіряти значення ФР на належність інтервалу. По-третє, якщо виникає потреба за значенням «гена» визначити значення відповідної йому ознаки, то попередньо закодовану інформацію розкодовують, щоб отримати числові або нечислові дані для виводу результатів, на що також витрачається процесорний час.

Тому пропонуємо інший метод представлення ФР: усі дані, необхідні для складання розкладу, представити у кодованому десятковому вигляді. Це дещо спростить кодування, адже, по-перше, не потрібно буде кодувати інформацію нечислових даних в числові, числові дані в бінарні, бінарні дані в код Грея, а, по-друге, потім усі ці дії проводити у зворотному порядку.

Початкова інформація про викладачів, групи, заняття та аудиторний фонд зберігається у таблицях зовнішньої бази даних і кодується автоматично при вводі. Отже, кодування частин ФР буде наступним: на кодування першої частини фрагмента відводиться 1 знак (дні тижня лежать в діапазоні [1;5]), другої – 1 знак (кількість пар – [1;5]), третьої – 6 знаків (20 груп, 40 викладачів, 80 предметів – [1..20;1..40;1..80]), четвертої – 3 знаки (аудиторії – [1;100]). При такому кодуванні ФР буде містити в собі 11 знаків, а хромосома складатиметься з $500 \cdot 11 = 5500$ знаків. Тобто десяткове кодування дає можливість скоротити наш масив приблизно в 3 рази. Наприклад, вищезображений ФР в десятковому кодуванні буде таким: 2 3 18 24 46 097, де 2 – день тижня (вівторок); 3 – номер пари (третья пара); 18 – номер групи, 24 – номер викладача, 46 – номер предмета; 097 – номер аудиторії.

Початкова популяція та формування цільової функції

Початкова популяція – це сукупність хромосом, яка, зазвичай, складає декілька десятків хромосом. Як було зазначено вище, для отримання нового кращого

покоління з існуючої популяції відбирають хромосоми, які будуть брати участь у подальшому схрещуванні. Існує декілька способів відбору хромосом. Всі вони базуються на тому чи іншому підході до визначення придатності хромосоми до схрещування. Ця придатність визначається за допомогою фітнес-функції або цільової функції (ЦФ) і безпосередньо впливає на ймовірність участі хромосоми в генерації розв'язків-нащадків:

$$\sum_{t \in T} \sum_{p=1}^P w(p) \cdot F(p, t) \rightarrow \max, \quad (1)$$

де $w(p)$ – ваговий коефіцієнт статусу викладача (p – номер викладача, $p = 1, \dots, P$); $F(p, t)$ – булева змінна (t – номер робочого дня тижня, $t \in T$ (T – множина номерів робочих днів для групи k)), яку можна знайти із обмеження виду:

$$1 \leq Q(p, t) + M \cdot F(p, t) \leq M \quad \forall t \in T; \quad \forall p = 1, \dots, P, \quad (2)$$

де M – довільне додатне досить велике число і

$$Q(p, t) = \sum_{k=1}^G \sum_{s=1}^S y(p, k, s) \cdot x(t, k, s), \quad (3)$$

$$\text{де } y(p, k, s) = \begin{cases} 1, & \text{якщо в день } t \text{ у групі } k \text{ проходить заняття } s; \\ 0, & \text{в іншому випадку;} \end{cases}$$

$$x(t, k, s) = \begin{cases} 1, & \text{якщо викладач } p \text{ веде заняття } s \text{ у групі } k; \\ 0, & \text{в іншому випадку.} \end{cases}$$

Із формули (2) випливає, що якщо $Q(p, t) = 0$, то $F(p, t) = 1$, і, якщо $Q(p, t) > 0$, то $F(p, t) = 0$.

Алгоритм спрямований на знаходження розв'язку з найбільшим значенням ЦФ, а ЦФ, в свою чергу, спрямована на оптимізацію робочого часу викладача.

Особливості реалізації генетичних операторів

Сам процес відбору та репродукції хромосом здійснюється так: з поточної популяції випадково вибирається пара хромосом і з них визначається краща (таких хромосом буде значно менше загальної кількості хромосом в популяції), причому функція відбору вказує на хромосому лише один раз, що дає можливість замінювати найгірші хромосоми на кращі в наступній популяції. Щоб отримати нову популяцію, відібрані хромосоми попарно схрещують між собою. Наш фрагмент розкладу представлений 4 частинами, кожна з яких – неподільна, але це не заважає нам точки кросоверу розміщувати не за границями фрагментів, а саме між частинами фрагмента. В даній задачі пропонується використати однорідний кросовер [10]: хромосоми розбивають на частини, причому фрагмент теж розбивається на частини, і кожна ділянка хромосоми першого батька має 50-відсотковий шанс обмінятися з відповідною ділянкою хромосоми другого батька. Наприклад:

| | | | | | | |
|----------|----------|----------|-----------|-----------|-----------|------------|
| батьки: | 2 | 3 | 18 | 24 | 46 | 097 |
| | 5 | 4 | 35 | 98 | 42 | 020 |
| нащадки: | 2 | 4 | 35 | 98 | 42 | 097 |
| | 5 | 3 | 18 | 24 | 46 | 020. |

Після кросоверу до популяції застосовують оператор мутації. Зазвичай, спочатку задається ймовірність мутації і алгоритм її здійснення. В бінарному

випадку оператор мутації полягає в інвертуванні символів у випадково вибраних позиціях. Якщо працюємо з десятковими числами, то випадково вибране число замінюється на будь-яке відмінне від нього:

до мутації: 2 3 18 24 46 097
після мутації: 5 3 18 24 46 028.

У прикладі показано, що зміни відбулися в 1-й (число вибрано з відрізка [1;5]) і 4-й (число вибрано з відрізка [1;100]) частинах.

Ймовірність мутації залежить від самої задачі, але, зазвичай, має досить мале значення (від $\approx 0,001$ до $\approx 0,01$ [11], [12]). Проте варто зазначити, що ймовірність мутації не повинна бути константою, адже існують випадки, коли схрещування не дає покращення, тоді мутація може допомогти змінити ситуацію [13]. Після реалізації мутації цільова функція перераховується.

Таким чином, помістивши початкову популяцію у створене штучне середовище і реалізувавши процеси селекції, кросоверу і мутації, ми одержимо ітераційний алгоритм пошуку оптимального розв'язку. Розглянутий алгоритм є не лише стійким до локальних мінімумів, але й забезпечує відносно швидкий пошук оптимального розв'язку.

Експериментальна верифікація

Під час практичної реалізації системи особлива увага зверталась на задачу написання «ядра» системи – методам розв'язання задачі. Оскільки не ставилася задача написати повнофункціональний комерційний продукт, інтерфейсна частина була написана для цілей тестування ядра і визначення меж застосовності алгоритмів, тому включає мінімум функціональних можливостей. Методи розв'язання написані з використанням об'єктно-орієнтованих технологій, що дозволить в майбутньому легко інкапсулювати їх в подальші модифікації системи, не порушуючи цілісності взаємодії різних алгоритмів.

Алгоритм розв'язання задачі тестувався на різних вибірках вихідних даних. У ролі тестових початкових даних були використані як реальні дані про групи, викладачів і предмети денної форми навчання Черкаського державного технологічного університету на 2005/2006 навчальний рік, так і випадково сформовані початкові дані (дисциплінам випадковим чином призначали аудиторії для проведення занять). В середньому проводилося від 5 до 10 тестів для кожної розмірності початкових даних. В результаті отримано дані, наведені в табл. 1.

Таблиця 1 – Залежність часу розв'язання задачі складання розкладу від розмірності задачі

| Розмір задачі – (кількість пар)*(кількість груп)*(кількість днів тижня) | Час розв'язання (хв.) | | |
|---|-----------------------|------|----------|
| | max | min | середній |
| 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 |

На рис. 2 зображений графік залежності середнього часу розв'язання задачі від розмірності задачі (кількості пар на тиждень і кількості груп).

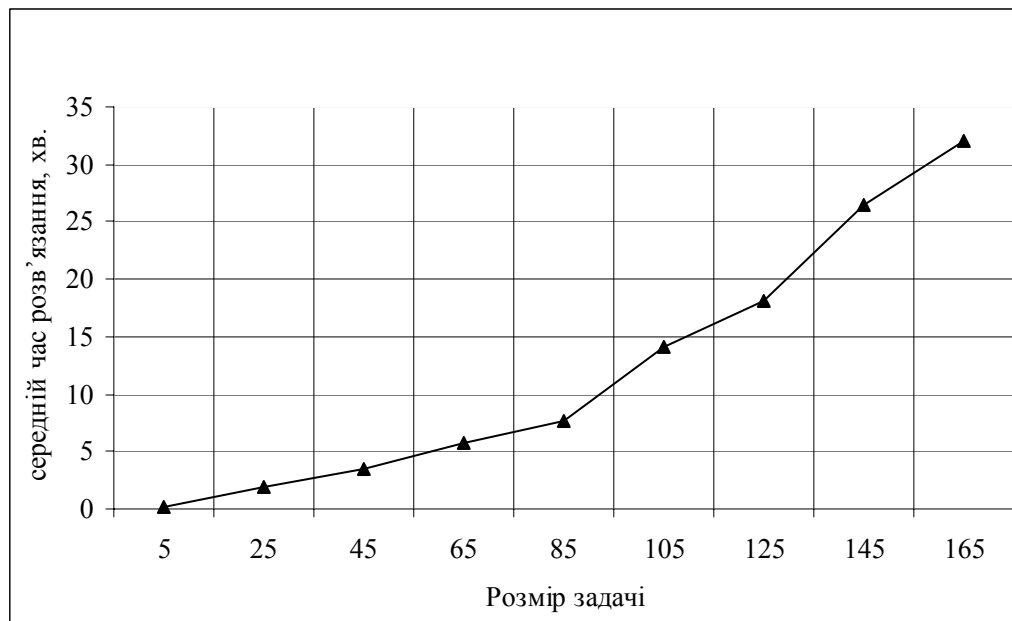


Рисунок 2 – Часова динаміка процесу розв'язання задачі

Висновки

Аналізуючи отримані дані, можна зробити такі висновки про функціональні можливості алгоритму розв'язання задачі, математичної моделі і області застосування.

По-перше, використана математична модель містить в собі значну кількість обмежень. По-друге, значно зростає час розв'язання задачі при збільшенні розмірності вихідних даних. Це відбувається через різке збільшення кількості обмежень в моделі, внаслідок чого зростає розмірність масивів і відповідно час розв'язання задачі. Але у разі використання інших методів оптимальність (або досягнення глобального максимуму) алгоритму може бути доведена лише повним перебором усіх можливих варіантів (ясно, що в цьому випадку час роботи алгоритму буде дуже великим), тому ітераційний пошук припиняється після досягнення певного максимального (не можна вказати, локального чи глобального) значення. Розв'язки, одержані за допомогою таких алгоритмів, можуть бути близькими до оптимальних, але не оптимальними. У цьому випадку для досягнення глобального максимуму можна використовувати представлений у роботі еволюційний метод, оскільки оптимум може бути знайдений за декілька ітерацій алгоритму. По-третє, математично формалізована задача охоплює лише задачу складання розкладу для студентів без урахування переходів між корпусами (ця вимога буде врахована в майбутньому) і, можливо, виконання інших вимог. Врахування додаткових обмежень збільшить кількість обмежень задачі, що негативно вплине на швидкість роботи алгоритму пошуку розв'язку.

Запропонована цільова функція не є єдино можливою. Введення інших цільових функцій не змінить обмеження математичної моделі і методів розв'язання задачі, але може істотно вплинути на результати розрахунків.

Загалом, запропонована еволюційна технологія дозволяє знайти оптимальний розв'язок задачі складання розкладів за декілька ітерацій, що істотно менше у порівнянні з відомими методами.

Література

1. Ерунов В.П., Морковин И.И. Формирование оптимального расписания учебных занятий в вузе // Вестник ОГУ. – 2001. – № 3. – С. 55-63.
2. Коффман Э.Г. Теория расписаний и вычислительные машины. – М.: Наука, 1984. – 335 с.
3. Губаев А., Бекенская Ю. Справедливая зарплата // The Chief (Шеф). – 2005. – № 3 (38). – С. 34-41.
4. Лагоша Б.А., Петропавловская А.В. Комплекс моделей и методов оптимизации расписания занятий в вузе // Экономика и математические методы. – 1993. – № 4. – С. 48-56.
5. Thompson J., Dowsland K. Variants of simulated annealing for the examination timetabling problem // Annals of Operational Research. – 1996. – Vol. 63.
6. Burke E.K., Elliman D.G., Weare R.F. A University Timetabling System Based on Graph Colouring and Constraint Manipulation // Journal of Research on Computing in Education. – 1993. – № 28. – P. 46-62.
7. Рубальская О.Н. Автоматизированные системы составления учебных расписаний. – М.: НИИВО, 2001. – 68 с.
8. Muller T. Some Novel Approaches to Lecture Timetabling // Proc. of the 4-th Workshop of Constraint Programming for Decision and Control "CPDC'2002". – Gliwice. – 2002. – P. 352-369.
9. Стариков А. Генетические алгоритмы – математический аппарат. – <http://www.basegroup.ru/genetic/math.htm>.
10. Syswerda G. Uniform crossover in genetic algorithms // Proceedings of the Third International Conference on Genetic Algorithms. – San Mateo, CA: Morgan Kaufmann Publishers, 1989. – P. 2-9.
11. Schaffler J.D., Caruana R.A., Escherman L.J., Das R. A study of control parameters affecting online performance of genetic algorithm for function optimisation // Proceedings of the Third International Conference on Genetic Algorithms and their Applications. – San Mateo, CA: Morgan Kaufmann Publishers, 1989. – P. 51-60.
12. Grefenstette J.J. Optimization of control parameters for genetic algorithms // IEEE Transaction on systems, man and cybernetics. SMC-16 (1). – 1986. – P. 122-128.
13. Davis L. Adapting operator probabilities in genetic algorithms // Proceedings of the Third International Conference on Genetic Algorithms. – La Jolla, CA: Morgan Kaufmann Publishers, 1989. – P. 60-69.

А.Н. Бойко

Эволюционная технология решения задачи составления расписаний учебных занятий

В статье изложено обоснование необходимости создания автоматизированной системы составления расписания занятий и аналитический обзор классических методов решения задачи составления расписаний. В качестве предметной области рассмотрено высшее учебное заведение и проблемы составления расписания учебных занятий студентов. Большое внимание уделено использованию эволюционных методов для решения задачи составления расписаний и исследованию результатов работы созданной системы.

А.М. Вуко

There is a necessity basis of automated system for scheduling creation and state-of-the-art review of classic methods for solving of scheduling task suggested. As subject there is higher school and problem of scheduling for students. The attention is spared to the use of evolutionary techniques for the problem decision of scheduling and to experimental work research of the created system.

Стаття надійшла до редакції 27.06.2006.