УДК 519.68

# Один із варіантів методу *m*-паралельного блочного пошуку записів і його ефективність

Володимир Лісовець[1], Григорій Цегелик[2]

[1] Львівський національний університет імені Івана Франка, вул. Університетська, 1, Львів, 79000, e-mail: kafmmsep@franko.lviv.ua
[2] д. ф.-м. н., професор, Львівський національний університет імені Івана Франка, вул. Університетська, 1, Львів, 79000, e-mail: kafmmsep@franko.lviv.ua

*Запропоновано один із варіантів методу **m**-паралельного блочного пошуку записів у файлах баз даних, орієнтований на використання в багатопроцесорних ЕОМ. Досліджено його ефективність для відомих законів розподілу ймовірностей звертання до записів (рівномірного, «бінарного», Зіпфа й узагальненого, частковим випадком якого є розподіл, який наближено задовольняє правило «80-20»). За критерій ефективності прийнято математичне сподівання кількості паралельних порівнянь, необхідних для пошуку запису у файлі. Побудовано оптимальні схеми методу, тобто схеми, за яких математичне сподівання досягає мінімуму. Зі збільшенням кількості процесорів у **k** разів ефективність запропонованого варіанта методу зростає в **k** разів, порівняно з ефективністю звичайного блочного методу, для всіх розглянутих законів розподілу ймовірностей звертання до записів, окрім «бінарного».*

**Ключові слова:** багатопроцесорні системи, *m*-паралельний пошук, блочний пошук, бази даних.

**Вступ.** Застосування паралельних обчислювальних систем є стратегічним напрямком розвитку обчислювальної техніки. Це викликано обмеженістю максимально можливої швидкодії звичайних послідовних ЕОМ, а також наявністю обчислювальних задач, для розв’язування яких можливості існуючих засобів обчислювальної техніки недостатні.

Проблема створення високопродуктивних обчислювальних систем належить до переліку найскладніших науково-технічних задач. Організація паралельних обчислень здійснюється, в основному, за рахунок уведення надлишкових функціональних пристроїв (декількох процесорів). Якщо здійснити поділ алгоритмів, які застосовуються, на інформаційно-незалежні частини й організувати виконання кожної частини обчислень на різних процесорах, то можна прискорити процес обчислень. Такий підхід дозволяє виконувати необхідні обчислення з меншими затратами часу. Одержання максимально можливого прискорення обмежується тільки кількістю наявних процесорів і «незалежних» частин алгоритму.

Однак слід відзначити, що на даний час застосування паралельних обчислень не одержало настільки широкого поширення, як це очікувалося багатьма дослідниками. Однією з можливих причин такої ситуації була донедавна висока вартість високопродуктивних багатопроцесорних систем. Кардинальні зміни відбулися з впровадженням порівняно дешевих багатоядерних процесорів, які вже набули масового застосування. Інша причина стримування поширення паралельних систем полягає в тому, що для проведення обчислень необхідно змінити традиційну послідовну технологію розв'язування задач на ЕОМ на паралельну.

Завдяки високій надійності та продуктивності багатопроцесорні ЕОМ широко використовують для підтримки й організації великих баз даних (БД). Під час розв'язування різноманітних задач із використанням БД основний акцент переноситься з процедур обробки інформації на процедури організації збереження та пошуку інформації в них. Тому продуктивність обчислювальних систем, орієнтованих на роботу з великими БД, значною мірою визначається ефективністю методів паралельного пошуку інформації в БД.

У працях [2, 3] розглянуто два паралельних методи пошуку інформації у файлах БД: метод $m$-паралельного послідовного перегляду та метод $m$-паралельного блочного пошуку. Вивчено ефективність цих методів для різних законів розподілу ймовірностей звертання до записів таких, як: рівномірний, «бінарний», Зіпфа й узагальнений, частковим випадком якого є розподіл, що наближено задовольняє правило «80-20» [1, 4, 8].

У роботі запропоновано інший варіант методу $m$-паралельного блочного пошуку записів у файлах БД, проведено дослідження його ефективності для різних законів розподілу ймовірностей звертання до записів. За критерій ефективності прийнято математичне сподівання кількості паралельних порівнянь, необхідних для пошуку запису у файлі. Побудовано оптимальні схеми методу, за реалізації яких математичне сподівання досягає мінімуму.

Дослідження ефективності $m$-паралельного блочного методу проведено для рівномірного розподілу ймовірностей звертання до записів і таких законів нерівномірного розподілу ймовірностей:

- «бінарний» розподіл

$$p_i = \frac{1}{2^i}, \quad i = \overline{1, N-1}, \qquad p_N = \frac{1}{2^{N-1}},$$

де $p_i$ — ймовірність звертання до $i$-го запису, $N$ — кількість записів у файлі;

- закон Зіпфа

$$p_i = \frac{1}{iH_N}, \quad i = \overline{1, N}, \quad H_N = \sum_{k=1}^{N} \frac{1}{k};$$

- узагальнений закон розподілу

$$p_i = \frac{1}{i^c H_N^{(c)}}, \quad i = \overline{1, N}, \quad H_N^{(c)} = \sum_{k=1}^{N} \frac{1}{k^c},$$

де $c$ ($0 < c < 1$) — довільний параметр.

Слід зауважити, що у випадку однопроцесорних ЕОМ ефективність методів пошуку для різних законів розподілу ймовірностей звертання до записів, а також порівняння методів за ефективністю проведено у працях [5-7, 9-11]. Деякі часткові випадки ефективності методів розглянуто в дослідженнях [1, 4].

## 1. Варіант методу *m*-паралельного блочного пошуку та його ефективність

Розглянемо впорядкований файл, який містить $N$ записів. Приймаємо, що записи файла поділено на $nm$ блоків, у кожному з яких є $sm$ записів. Зауважимо, якщо дане розбиття для цілих значень $n, m, s$ не можна реалізувати, то доповнюємо файл пустими записами. Тоді сумарна кількість записів у файлі буде $N = snm^2$. Пошук запису у файлі здійснюється так. Спочатку знаходимо блок, який містить шуканий запис, використовуючи метод $m$-паралельного послідовного перегляду серед останніх елементів блоків. Після цього з допомогою методу $m$-паралельного послідовного перегляду продовжуємо пошук у локалізованому блоці.

Дослідимо ефективність цього методу для різних законів розподілу ймовірностей звертання до записів.

Нехай $p_i$ — ймовірність звертання до $i$-го запису файла. Математичне сподівання $E$ кількості паралельних порівнянь, необхідних для пошуку запису у файлі, подамо як суму математичного сподівання кількості паралельних порівнянь, що потрібні для локалізації блоку записів, і математичного сподівання кількості паралельних порівнянь, потрібних для пошуку запису в локалізованому блоці. Тоді

$$E = \sum_{k=1}^{n}\sum_{l=1}^{m} k\left(\sum_{i=1}^{s}\sum_{j=1}^{m} p_{(k-1)m^2 s-(l-1)ms+(i-1)m+j}\right) + \sum_{k=1}^{nm}\sum_{i=1}^{s}\sum_{j=1}^{m} i p_{(k-1)ms+(i-1)m+j} .$$

Знайдемо явний вираз для $E$ у випадку різних законів розподілу ймовірностей звертання до записів і дослідимо залежність $E$ від зміни закону розподілу ймовірностей. Побудуємо оптимальні схеми $m$-паралельного блочного пошуку для розглянутих законів розподілу ймовірностей звертання до записів.

**1.1. Рівномірний розподіл.** Якщо розподіл імовірностей звертання до записів є рівномірний, то для $E$ одержуємо вираз

$$E = \frac{1}{2}(n+1) + \frac{1}{2}(s+1) .$$

Функція $E$ досягає мінімуму, якщо $n = s = \sqrt{N}/m$ . Тоді $\min E = \sqrt{N}/m + 1$ .

**1.2. «Бінарний» розподіл.** Нехай імовірності звертання до записів задовольняють «бінарний» розподіл. Тоді для $E$ матимемо формулу [8]

$$E = \frac{s}{2^N} + \left(\frac{2^{m^2 s}}{2^{m^2 s}-1} + \frac{2^m}{2^m-1} - \frac{s}{2^{ms}-1}\right)\left(1-2^{-N}\right).$$

Нехтуючи нескінченно малою величиною $2^{-N}$, із достатньо високою точністю можемо прийняти

$$E = \frac{2^{m^2 s}-s}{2^{m^2 s}-1} + \frac{2^m}{2^m-1} - \frac{s}{2^{ms}-1} .$$

Оскільки $m$ фіксований параметр, то функція $E$ залежить лише від $s$. Обчислимо похідну від функції $E$ по $s$. Одержимо

$$\frac{dE}{ds}=\frac{2^{ms}(sm\ln 2-1)+1}{(2^{ms}-1)^2}-\frac{2^{m^2s}m^2\ln 2}{(2^{m^2s}-1)^2}$$

або

$$\frac{dE}{ds}=\left(2^{ms}-1\right)^{-2}\left(2^{m^2s}-1\right)^{-2}\left\{2^{m^2s+ms}\left[\left(2^{m^2s}-2\right)\left(sm\ln 2-1\right)-\right.\right.$$

$$\left.\left.-\left(2^{ms}-2\right)m^2\ln 2\right]+\left[2^{m^2s}\left(2^{m^2s}-m^2\ln 2-2\right)+2^{ms}\left(sm\ln 2-1\right)+1\right]\right\}.$$

Покажемо, що для $m\geq 2$, $s\geq 2$ виконується умова $E'>0$, тобто, якщо $s\geq 2$, $m\geq 2$, то функція $E$ монотонно зростає. Справді, якщо $s\geq 2$, $m\geq 2$, то завжди виконуються нерівності

$$2^{m^2s}-m^2\ln 2-2>0\,,\qquad sm\ln 2-1>1\,.$$

Тому достатньо довести, що для $s\geq 2$, $m\geq 2$ справджується така нерівність

$$(2^{m^2s}-2)(sm\ln 2-1)-(2^{ms}-2)m^2\ln 2>0\,.$$

Із очевидної нерівності $2^{m\,(m-1)}>m^2\ln 2$, яка виконується для $m\geq 2$, $s\geq 2$, отримаємо

$$2^{ms(m-1)}>m^2\ln 2\quad \text{або}\quad 2^{m^2s-ms}-m^2\ln 2>0\,.$$

Якщо $m\geq 2$, $s\geq 2$, то $(2-2m^2\ln 2)/2^{ms}<0$. Тоді

$$2^{m^2s-ms}-m^2\ln 2>\frac{2-2m^2\ln 2}{2^{ms}}\,.$$

Звідси можна одержати, що $2^{m^2s}-2>(2^{ms}-2)m^2\ln 2$, або

$$(2^{m^2s}-2)(sm\ln 2-1)>(2^{ms}-2)m^2\ln 2\,.$$

Що і треба було довести. Тому $E'>0$, якщо $s\geq 2$, $m\geq 2$.

Отже, якщо $s\geq 2$, $m\geq 2$, то функція $E$ монотонно зростає. Оскільки $s$ — ціле, то в області $s\geq 1$ функція досягає мінімального значення при $s=1$ або $s=2$. Обчислимо значення $E$ у цих точках

$$E(2)=\frac{2^{2m^2}}{2^{2m^2}-1}+\frac{2^m}{2^m-1}-\frac{2}{2^{2m}-1}=1+\frac{1}{2^{2m^2}-1}+\frac{2^m}{2^m-1}-\frac{1}{2^m-1}+\frac{1}{2^m+1}\,,$$

$$E(1)=\frac{2^{2m^2}}{2^{m^2}-1}+\frac{2^m}{2^m-1}-\frac{2}{2^m-1}=1+\frac{1}{2^{m^2}-1}+\frac{2^m}{2^m-1}-\frac{1}{2^m-1}\,.$$

Оскільки $E(2)-E(1)=\dfrac{1}{2^{2m^2}-1}+\dfrac{2^m}{2^m+1}-\dfrac{1}{2^{m^2}-1}>0$ для довільних $m\geq 2$, то $E(2)>E(1)$. Отже, в області $s\geq 1$ функція $E$ досягає мінімуму, якщо $s=1$.

**1.3. Закон Зіпфа.** Нехай імовірності звертання до записів задовольняють закон Зіпфа. Тоді для $E$ одержуємо вираз

$$E=\frac{1}{H_N}\Big[(n+2)H_N+sS_{ms}(nm)-S_{m^2s}(n)-S_m(nms)\Big],$$

де

$$S_{ms}(nm)=\sum_{k=1}^{nm}H_{kms}\ ,\quad S_{m^2s}(n)=\sum_{k=1}^{n}H_{km^2s}\ ,\quad S_m(nms)=\sum_{k=1}^{nms}H_{km}\ .$$

Використаємо для сум $S_{ms}(nm)$, $S_{m^2s}(n)$ та $S_m(nms)$ апроксимації [8]

$$S_{ms}(nm)\approx nm\big(H_N-1\big)+\frac{1}{2}\ln(nm)+C_1,$$

$$S_{m^2s}(n)\approx n\big(H_N-1\big)+\frac{1}{2}\ln n+C_1,$$

$$S_m(nms)\approx nms\big(H_N-1\big)+\frac{1}{2}\ln(nms)+C_1,$$

де $C_1$ = 0,5ln2π. Тоді з достатньо високою точністю можемо прийняти

$$E=\frac{1}{H_N}\left[2H_N+n+\frac{1}{2}s\ln(nm)-\frac{1}{2}\ln(nms)-\frac{1}{2}\ln n+(s-2)C_1\right],$$

або

$$E=\frac{1}{H_N}\left[2H_N+n+\frac{1}{2}\frac{N}{m^2n}\ln(nm)-\frac{1}{2}\ln n-\frac{1}{2}\ln\left(\frac{N}{m}\right)+C_1\left(\frac{N}{m^2n}-2\right)\right].$$

Оскільки

$$\frac{dE}{dn}=\frac{1}{H_N}\left\{1+\frac{N}{2m^2n^2}\big[1-\ln(nm)-2C_1\big]-\frac{1}{2n}\right\},$$

то для знаходження значення параметра $n$, за якого $E$ досягає мінімуму, одержуємо рівняння

$$2n^2-n=\frac{N}{m^2}(\ln n+\ln m+2C_1-1)\ .$$

**1.4. Узагальнений закон.** Нехай імовірності звертання до записів задовольняють узагальнений закон розподілу. Тоді для визначення математичного сподівання $E$ маємо формулу [8]

$$E=\frac{1}{H_N^{(c)}}\Big[(n+2)H_N^{(c)}+sS_{ms}^{(c)}(nm)-S_{m^2s}^{(c)}(n)-S_m^{(c)}(nms)\Big],$$

де

$$S_{ms}^{(c)}(nm)=\sum_{k=1}^{nm}H_{kms}^{(c)}\ ,\quad S_{m^2s}^{(c)}(n)=\sum_{k=1}^{n}H_{km^2s}^{(c)}\ ,\quad S_m^{(c)}(nms)=\sum_{k=1}^{nms}H_{km}^{(c)}\ .$$

Використовуючи для $S_{ms}^{(c)}(nm)$, $S_{m^2s}^{(c)}(n)$ та $S_{m}^{(c)}(nms)$ відповідні апроксимації [8]

$$S_{ms}^{(c)}(nm) \approx nmH_N^{(c)} + \frac{N^{1-c}}{1-c}\left[\frac{c-1}{2-c}nm + \frac{\alpha^{(c)}(nm)}{(nm)^{1-c}}\right],$$

$$S_{m^2s}^{(c)}(n) \approx nH_N^{(c)} + \frac{N^{1-c}}{1-c}\left[\frac{c-1}{2-c}n + \frac{\alpha^{(c)}(n)}{n^{1-c}}\right],$$

$$S_{m}^{(c)}(nms) \approx nmsH_N^{(c)} + \frac{N^{1-c}}{1-c}\left[\frac{c-1}{2-c}nms + \frac{\alpha^{(c)}(nms)}{(nms)^{1-c}}\right],$$

де

$$\alpha^{(c)}(n) = H_n^{(c-1)} - \frac{1}{2-c}n^{2-c}, \qquad \alpha^{(c)}(nm) = H_{nm}^{(c-1)} - \frac{1}{2-c}(nm)^{2-c},$$

$$\alpha^{(c)}(nms) = H_{nms}^{(c-1)} - \frac{1}{2-c}(nms)^{2-c}$$

є повільно зростаючими функціями, з достатньо високою точністю можемо прийняти, що

$$E = \frac{1}{H_N^{(c)}}\left\{2H_N^{(c)} + \frac{N^{1-c}}{1-c}\left[\frac{1-c}{2-c}n + \frac{s\alpha^{(c)}(nm)}{(nm)^{1-c}} - \frac{\alpha^{(c)}(n)}{n^{1-c}} - \frac{\alpha^{(c)}(nms)}{(nms)^{1-c}}\right]\right\}$$

або

$$E = \frac{1}{H_N^{(c)}}\left\{2H_N^{(c)} - \frac{N^{1-c}}{1-c}\left[\frac{1-c}{2-c}n + \frac{N}{m^2 n}\frac{\alpha^{(c)}(nm)}{(nm)^{1-c}} - \frac{\alpha^{(c)}(n)}{n^{1-c}} - \frac{\alpha^{(c)}(N/m)}{(N/m)^{1-c}}\right]\right\}.$$

Обчислимо похідну від функції $E$ по $n$, підставляючи замість похідних від $\alpha^{(c)}(n)$ і $\alpha^{(c)}(nm)$ відповідно різниці $\alpha^{(c)}(n+1) - \alpha^{(c)}(n)$ і $\alpha^{(c)}(nm+1) - \alpha^{(c)}(nm)$. Одержимо

$$\frac{dE}{dn} \approx \frac{1}{H_N^{(c)}}\left\langle \frac{N^{1-c}}{1-c}\left\{\frac{1-c}{2-c} - \frac{N}{(mn)^{3-c}}\left[(n+2-c)\alpha^{(c)}(nm) - n\alpha^{(c)}(nm+1)\right] - \right.\right.$$

$$\left.\left. - \frac{1}{n^{2-c}}\left[n\alpha^{(c)}(n+1) - (n+1-c)\alpha^{(c)}(n)\right]\right\}\right\rangle.$$

Тому, для наближеного обчислення значень параметра $n$, за яких функція $E$ досягає мінімуму, маємо рівняння

$$n^{3-c} + \frac{2-c}{1-c}n\left[(n+1-c)\alpha^{(c)}(n) - n\alpha^{(c)}(n+1)\right] =$$

$$= \frac{N}{m^{3-c}}\frac{2-c}{1-c}\left[(n+2-c)\alpha^{(c)}(nm) - n\alpha^{(c)}(nm+1)\right].$$

**1.5. Порівняння результатів.** Оптимальні значення параметра $n$, за яких математичне сподівання кількості паралельних порівнянь, необхідних для пошуку запису у файлі, досягає мінімуму, й оптимальні значення математичного сподівання для $N = 10^6$, деяких $m$ і різних законів розподілу ймовірностей звертання до записів приведені відповідно в табл. 1 і 2.

*Таблиця 1*

Оптимальні значення параметра $n$ для різних законів розподілу ймовірностей і різної кількості процесорів

| $m$ | Рівномірний | Узагальнений | | | | Зіпфа | «Бінарний» |
|---|---|---|---|---|---|---|---|
| | | $c$ = 0,2 | $c$ = 0,4 | $c$ = 0,6 | $c$ = 0,8 | | |
| 1 | 1000 | 1060 | 1150 | 1297 | 1556 | 2058 | 1000000 |
| 2 | 500 | 530 | 575 | 648 | 778 | 1029 | 250000 |
| 4 | 250 | 265 | 288 | 324 | 389 | 515 | 62500 |
| 5 | 200 | 212 | 230 | 259 | 311 | 412 | 40000 |
| 10 | 100 | 106 | 115 | 130 | 156 | 206 | 10000 |
| 20 | 50 | 53 | 58 | 65 | 78 | 103 | 2500 |
| 40 | 25 | 27 | 29 | 32 | 39 | 52 | 625 |
| 50 | 20 | 21 | 23 | 26 | 31 | 41 | 400 |
| 100 | 10 | 11 | 12 | 13 | 16 | 21 | 100 |

*Таблиця 2*

Оптимальні значення математичного сподівання кількості паралельних порівнянь для різних законів розподілу ймовірностей і різної кількості процесорів

| $m$ | Рівномірний | Узагальнений | | | | Зіпфа | «Бінарний» |
|---|---|---|---|---|---|---|---|
| | | $c$ = 0,2 | $c$ = 0,4 | $c$ = 0,6 | $c$ = 0,8 | | |
| 1 | 1001 | 943,44 | 865,01 | 748,60 | 561,44 | 303,93 | 3,00 |
| 2 | 501 | 472,22 | 433,01 | 374,81 | 281,26 | 152,58 | 2,07 |
| 4 | 251 | 236,61 | 217,01 | 187,92 | 141,18 | 76,92 | 2,00 |
| 5 | 201 | 189,49 | 173,81 | 150,54 | 113,17 | 61,80 | 2,00 |
| 10 | 101 | 95,25 | 87,41 | 75,79 | 57,15 | 31,57 | 2,00 |
| 20 | 51 | 48,13 | 44,22 | 38,43 | 29,16 | 16,48 | 2,00 |
| 40 | 26 | 24,57 | 22,62 | 19,75 | 15,18 | 8,95 | 2,00 |
| 50 | 21 | 19,86 | 18,30 | 16,02 | 12,39 | 7,46 | 2,00 |
| 100 | 11 | 10,44 | 9,68 | 8,57 | 6,82 | 4,48 | 2,00 |

**Висновки.** У роботі запропоновано один із варіантів методу $m$-паралельного блочного пошуку записів у файлах баз даних, який орієнтований на використання в багатопроцесорних ЕОМ. Досліджено ефективність цього методу для різних законів розподілу ймовірностей звертання до записів: рівномірного, «бінарного», Зіпфа й узагальненого. За критерій ефективності прийнято математичне сподівання кількості паралельних порівнянь, необхідних для пошуку запису у файлі. Виведено співвідношення для оптимальних значень параметра $n$, при яких математичне сподівання досягає мінімуму.

Порівнюючи ефективність методу блочного пошуку [11] і запропонованого варіанта методу $m$-паралельного блочного пошуку, приходимо до висновку, що розпаралелювання методу блочного пошуку для всіх розглянутих законів розподілу ймовірностей звертання до записів, окрім «бінарного», суттєво підвищує ефективність. У випадку «бінарного» закону розподілу ймовірностей збільшення кількості процесорів не підвищує ефективності роботи. Якщо порівняти ефективність даного варіанта $m$-паралельного блочного пошуку з варіантом, запропонованим раніше в [3], то приходимо до такого висновку: якщо $m = 1$, то обидва варіанти методу однаково ефективні; але зі зростанням кількості процесорів ефективність запропонованого тут варіанта методу значно зростає, порівняно з ефективністю методу, запропонованого в праці [3].

## Література

[1] *Кнут Д.* Искусство программирования для ЭВМ.— М.: Изд. дом «Вильямс», 2000. — Т. 3: Сортировка и поиск. — 832 с.

[2] *Лісовець В. Я., Цегелик Г. Г.* Метод $m$-паралельного послідовного перегляду записів та його використання для пошуку інформації у послідовних файлах баз даних // Фізико-математичне моделювання та інформаційні технології. — 2007. — Вип. 5. — С. 109-119.

[3] *Лісовець В. Я., Цегелик Г. Г.* Метод $m$-паралельного блочного пошуку записів у файлах баз даних та його ефективність // Відбір та обробка інформації. — 2007. — Вип. 27(103). — С. 87-92.

[4] *Мартин Дж.* Организация баз данных в вычислительных системах. — М: Мир, 1980. — 644 с.

[5] *Мельничин А. В., Цегелик Г. Г.* Аналіз методів пошуку інформації в файлах баз даних для різних законів розподілу ймовірностей звертання до записів // Комп’ютерні технології друкарства. — 2006. — № 15. — С. 95-112.

[6] *Мельничин А. В., Цегелик Г. Г.* Ефективність методу двійкового пошуку інформації у файлах баз даних для різних законів розподілу ймовірностей звертання до записів // Вісн. Львів. ун-ту. Сер. прикл. матем. та інформ. — 2006. — Вип. 11. — С. 225-229.

[7] *Мельничин А. В., Цегелик Г. Г.* Методи пошуку інформації у файлах баз даних та їх ефективність для різних законів розподілу ймовірностей звертання до записів // Комп’ютерні технології друкарства. — 2006. — № 16. — С. 41-52.

[8] *Цегелик Г. Г.* Организация и поиск информации в базах данных. — Львов: Вища шк., 1987. — 176 с.

[9] *Цегелик Г. Г., Мельничин А. В.* Порівняльний аналіз ефективності методів пошуку інформації у файлах баз даних // Відбір і обробка інформації. — 2005. — № 23. — С. 135-142.

[10] *Цегелик Г., Мельничин А.* Метод пошуку інформації у файлах баз даних, який враховує розподіл імовірностей звертання до записів // Фізико-математичне моделювання та інформаційні технології. — 2006. — Вип. 4. — С. 169-176.

[11] *Філяк М. І., Цегелик Г. Г., Дороцька Х. С.* Порівняльний аналіз ефективності методу послідовного перегляду для різних законів розподілу ймовірностей звертання до записів // Вісник НУ «Львівська політехніка». «Інформаційні системи та мережі». — 2000. — № 406. — С. 226-231.

## One of the version of *m*-parallel block record browsing method and its effectiveness

Volodymyr Lisovets, Hryhorii Tsehelyk

*A version of* ***m****-parallel block record browsing method in a database file is proposed. The method is designed for application in multiprocessors of computers. The effectiveness of the method for different probability distribution of record request frequency (discrete uniform, binomial, Zipf and generalized, the partial case of which is the probability distribution approximately satisfying the rule «80-20») is investigated. The mathematical expectation of the parallel comparisons number needed for the search of a record in a file is taken as a criterion of effectiveness. The optimal schemes of the method i. e. the scheme in which the mathematical expectation attains its minimum are proposed. When the number of processors increases in* ***k*** *times the effectiveness of the proposed method increases in* ***k*** *times as compared with the effectiveness of the conventional block method for all examined probability distributions of record request frequency except for binomial one.*

## Вариант метода *m*-параллельного блочного поиска записей и его эффективность

Владимир Лисовец, Григорий Цегелик

*Предлагается вариант метода* ***m****-параллельного блочного поиска записей в файлах баз данных, ориентированный на использование в многопроцессорных ЭВМ. Исследуется эффективность этого метода для известных законов распределения вероятностей обращения к записям (равномерного, «бинарного», Зипфа и обобщенного, частным случаем которого является распределение, приближенно удовлетворяющее правило «80-20»). В качестве критерия эффективности принимается математическое ожидание количества параллельных сравнений, необходимых для поиска записи в файле. Построены оптимальные схемы метода, при которых математическое ожидание достигает минимума. С увеличением количества процессоров в* ***k*** *раз эффективность предложенного варианта метода возрастает в* ***k*** *раз по сравнению с эффективностью обычного блочного метода для всех рассмотренных законов распределения вероятностей обращения к записям за исключением «бинарного».*

Отримано 15.11.07