УДК 004.272.26

Побудова оптимальних стратегій вибору інформації у послідовних файлах баз даних за використання методу *m*-паралельного блочного пошуку

Григорій Цегелик[1], Володимир Лісовець[2]

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

Розглядається використання методу m-паралельного блочного пошуку для відшукання записів у впорядкованих файлах баз даних, які зберігаються у зовнішній пам'яті багатопроцесорної ЕОМ. Із використанням методу m-паралельного блочного пошуку будуються оптимальні стратегії пошуку для таких законів розподілу ймовірностей звертання до записів, як: рівномірний, «бінарний», Зіпфа й узагальнений, частковим випадком якого є розподіл, що наближено задовольняє правило «80 – 20». За критерій оптимальності взято математичне сподівання загального часу, необхідного для пошуку запису у файлі. Оптимальні стратегії пошуку записів у впорядкованих файлах із використанням розглянутого варіанта методу m-паралельного блочного пошуку доволі суттєво залежить від закону розподілу ймовірностей звертання до записів.

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

Вступ. Завдяки високій надійності та продуктивності багатопроцесорні ЕОМ широко використовуються 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.

Дослідження ефективності методів пошуку інформації в БД є доволі складна задача. Зазвичай, за критерій ефективності береться середня кількість порівнянь, необхідних для пошуку запису. На практиці це теоретичне середнє часто відрізняється від реальної середньої кількості порівнянь. Насамперед, це пов'язано з тим, що ймовірності звертання до записів у файлах БД підпорядковані нерівномірним законам розподілу: одні записи шукаються часто, інші вкрай рідко. У цій роботі приймаємо, що ми можемо визначити закон розподілу, згідно якого розподілені записи. Це ми можемо зробити, наприклад, маючи статистичну інформацію конкретної БД.

У працях [1-5] проведено аналіз методів *m*-паралельного послідовного перегляду та двох варіантів методу *m*-паралельного блочного пошуку записів для різних законів розподілу ймовірностей звертання до записів, де за критерій ефективності взято математичне сподівання кількості паралельних порівнянь, необхідних для пошуку запису у файлі. Зауважимо, що побудова оптимальних стратегій пошуку інформації у послідовних файлах за використання методів послідовного перегляду та блочного пошуку записів для однопроцесорних ЕОМ і різних законів розподілу ймовірностей звертання до записів розглянуті в монографії [6]. А у праці [4] побудовано оптимальні стратегії пошуку записів у послідовних файлах, які зберігаються у зовнішній пам'яті багатопроцесорної ЕОМ, використовуючи метод m-паралельного послідовного перегляду.

У цій роботі, використовуючи метод m-паралельного блочного пошуку, побудовано оптимальні стратегії пошуку записів у файлах, які зберігаються у зовнішній пам'яті багатопроцесорної ЕОМ.

Побудову оптимальних стратегій пошуку проведено для рівномірного закону розподілу ймовірностей звертання до записів і таких законів нерівномірного розподілу ймовірностей, як у роботах [1-5, 7-9]:

- «бінарний» розподіл,
- закон Зіпфа,
- узагальнений закон розподілу.

1. Побудова математичних моделей пошуку інформації та знаходження оптимальних розмірів блоків

Нехай послідовний упорядкований файл, який містить N записів, знаходиться у зовнішній пам'яті багатопроцесорної ЕОМ, до складу якої входять m процесорів, які працюють паралельно та мають спільне поле пам'яті. Нехай файл умовно розбито на n блоків по sm записів у кожному ($N = nsm$) і пошук потрібного запису відбувається так. Шляхом послідовного зчитування блоків записів в основну пам'ять і перегляду їх останніх m записів локалізуємо блок, який містить шуканий запис. Після цього пошук продовжуємо вже у локалізованому блоці з допомогою методу m-паралельного послідовного перегляду [1, 3].

Нехай $a_0 = b_0 + d_0ms$ — час зчитування блоку записів в основну пам'ять, де b_0, d_0 — деякі сталі; t_0 — час виконання операції m-паралельного послідовного перегляду записів в основній пам'яті; p_i — ймовірність звертання до i-го запису файла; E_t — математичне сподівання загального часу, необхідного для пошуку запису у файлі. Подамо E_t як суму математичного сподівання часу, необхідного для локалізації блоку, та математичного сподівання часу, необхідного для пошуку запису в локалізованому блоці. Тоді E_t виразимо формулою

$$E_t = \sum_{k=1}^{n}\left(a_0 + t_0\right)k\sum_{i=1}^{s}\sum_{j=1}^{m} p_{(k-1)ms+(i-1)m+j} + \sum_{k=1}^{n}\sum_{i=1}^{s} t_0 i \sum_{j=1}^{m} p_{(k-1)ms+(i-1)m+j}$$

або

$$E_t = \sum_{k=1}^{n}\sum_{i=1}^{s}\sum_{j=1}^{m}\left[a_0k + (k+i)t_0\right]p_{(k-1)ms+(i-1)m+j}\ .$$

Знайдемо явний вираз для E_t у випадку різних законів розподілу ймовірностей звертання до записів і визначимо, за яких значень параметрів n і s математичне сподівання досягає мінімуму. Для цього введемо позначення

$$E_1 = \sum_{k=1}^{n}\sum_{i=1}^{s}\sum_{j=1}^{m} i p_{(k-1)ms+(i-1)m+j}, \qquad E_2 = \sum_{k=1}^{n}\sum_{i=1}^{s}\sum_{j=1}^{m} k p_{(k-1)ms+(i-1)m+j}.$$

Тоді

$$E_t = \left(a_0 + t_0\right) E_2 + t_0 E_1 .$$

1.1. Рівномірний розподіл. Нехай розподіл ймовірностей є рівномірний, тобто $p_i = 1/N$, $i = \overline{1,N}$.

Тоді $E_1 = (s+1)/2$, $E_2 = (n+1)/2$, $E_t = \left[(n+1)\left(a_0 + t_0\right) + (s+1)t_0\right]/2$ або $E_t = \dfrac{1}{2}\left[\left(\dfrac{N}{ms}+1\right)\left(b_0 + d_0 ms\right) + \left(\dfrac{N}{ms} + s + 2\right)t_0\right]$.

Оскільки

$$\frac{dE_t}{ds} = \frac{1}{2}\left[-\frac{N}{ms^2}\left(b_0 + t_0\right) + d_0 m + t_0\right],$$

то для знаходження параметра s, за якого функція E_t досягає мінімуму (m вважаємо сталим), одержуємо

$$s = \left\{\frac{N\left(b_0 + t_0\right)}{d_0} \bigg/ \left[m\left(m + \frac{t_0}{d_0}\right)\right]\right\}^{1/2}.$$

Тоді $n = N/ms$.

1.2. «Бінарний» розподіл. Якщо ймовірності звертання до записів задовольняють «бінарний» розподіл, тобто $p_i = 1/2^i, i = \overline{1,\, N-1}$, $p_N = 1/2^{N-1}$, то, аналогічно як у [6],

$$E_1 = \frac{s}{2^N} + \left(\frac{2^m}{2^m - 1} - \frac{s}{2^{ms} - 1}\right)\left(1 - 2^{-N}\right), \qquad E_2 = \frac{2^{ms}}{2^{ms} - 1}\left(1 - 2^{-N}\right).$$

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

$$E_t = \frac{2^{ms}}{2^{ms} - 1}\left(a_0 + t_0\right) + \left(\frac{2^m}{2^m - 1} - \frac{s}{2^{ms} - 1}\right)t_0$$

або

$$E_t = \frac{2^{ms}}{2^{ms} - 1}\left(b_0 + d_0 ms\right) + \left(\frac{2^m}{2^m - 1} + \frac{2^{ms} - s}{2^{ms} - 1}\right)t_0 .$$

Оскільки

$$\frac{dE_t}{ds} = -\frac{2^{ms} m \ln 2}{\left(2^{ms}-1\right)^2}\left(b_0 + d_0 ms\right) + \frac{2^{ms}}{2^{ms}-1} d_0 m + \frac{2^{ms}\left[(s-1)m\ln 2 - 1\right]+1}{\left(2^{ms}-1\right)^2} t_0 ,$$

то для знаходження параметра s, за якого E_t приймає найменше значення, одержуємо рівняння

$$\left(2^{ms}-1\right)m + \left[(s-1)m\ln 2 - 1 + 2^{-ms}\right]\frac{t_0}{d_0} = m\left(\frac{b_0}{d_0} + ms\right)\ln 2 .$$

На множині $m \geq 2$, $s \geq 2$, m, $s \in N$, справджується умова

$$2^{ms}\left[(s-1)m\ln 2 - 1\right] + 1 > 0 .$$

Щоб вираз

$$2^{ms} d_0 m\left(2^{ms}-1\right) - 2^{ms} m \ln 2\left(b_0 + d_0 ms\right)$$

був додатний, повинна виконуватись умова

$$2^{ms} - 1 - \left(\frac{b_0}{d_0} + ms\right)\ln 2 > 0 .$$

Тому за використання цієї умови на множині $m \geq 2$, $s \geq 2$, m, $s \in N$, функція E_t набуває найменшого значення для $s = 2$. У цьому випадку

$$E_t = \frac{4^m}{4^m - 1}\left(b_0 + 2d_0 m\right) + \left(\frac{2^m}{2^m - 1} + \frac{4^m - 2}{4^m - 1}\right)t_0 .$$

1.3. Закон Зіпфа. Приймемо, що розподіл ймовірностей звертання до записів задовольняє закон Зіпфа, тобто $p_i = 1/iH_N$, $i = \overline{1,N}$, де $H_N = \sum_{k=1}^{N}\frac{1}{k}$ — часткова сума гармонічного ряду. Тоді, аналогічно як у монографії [6], одержуємо

$$E_1 = \frac{1}{H_N}\left[H_N + s\, S_{ms}(n) - S_m(sn)\right], \qquad E_2 = \frac{1}{H_N}\left[(n+1)H_N - S_{ms}(n)\right],$$

де

$$S_{ms}(n) = \sum_{k=1}^{n} H_{kms} , \qquad S_m(ns) = \sum_{k=1}^{ns} H_{km} .$$

Тому

$$E_t = \frac{1}{H_N}\left\{\left[(n+1)H_N - S_{ms}(n)\right]\left(a_0 + t_0\right) + \left[H_N + s\,S_{ms}(n) - S_m(sn)\right]t_0\right\} .$$

Використовуючи апроксимації $S_{ms}(n)$ і $S_m(ns)$, відповідно, виразами [9]

$$\overline{S}_{ms}(n)=n\left(H_N-1\right)+\frac{1}{2}\ln n+C_1,\qquad \overline{S}_m(ns)=ns\left(H_N-1\right)+\frac{1}{2}\ln(ns)+C_1,$$

де $C_1 = 0,5\ln 2\pi$, із достатньо високою точністю можемо прийняти

$$E_t=\frac{1}{H_N}\left\{\left(H_N+n-\frac{1}{2}\ln n-C_1\right)\left(a_0+t_0\right)+\left[H_N+(s-1)\left(\frac{1}{2}\ln n+C_1\right)-\frac{1}{2}\ln s\right]t_0\right\}$$

або

$$E_t=\frac{1}{H_N}\left\{\left(H_N+n-\frac{1}{2}\ln n-C_1\right)\left(b_0+t_0+\frac{Nd_0}{n}\right)+\right.$$
$$\left.+\left[H_N+\left(\frac{N}{mn}-1\right)\left(\frac{1}{2}\ln n+C_1\right)-\frac{1}{2}\ln\frac{N}{nm}\right]t_0\right\}.$$

Оскільки

$$\frac{dE_t}{dn}=\frac{1}{H_N}\left\{\left(1-\frac{1}{2n}\right)\left(b_0+t_0+\frac{Nd_0}{n}\right)-\frac{Nd_0}{n^2}\left(H_N+n-\frac{1}{2}\ln n-C_1\right)+\right.$$
$$\left.+\left[\frac{1}{2}\left(-\frac{N}{mn^2}\left(\ln n+2C_1\right)+\left(\frac{N}{mn}-1\right)\frac{1}{n}+\frac{1}{n}\right)\right]t_0\right\},$$

то для знаходження значення параметра n, за якого E_t досягає найменшого значення, одержуємо рівняння

$$n\left(2n-1\right)m\frac{b_0+t_0}{d_0}=N\left[2H_Nm+\left(\frac{t_0}{d_0}-m\right)\left(\ln n+2C_1-1\right)\right].$$

1.4. Узагальнений закон. Нехай розподіл імовірностей звертання до записів задовольняє узагальнений закон розподілу, тобто $p_i=1/i^cH_N^{(c)}, i=\overline{1,N}$, де c $(0<c<1)$ — будь-який параметр, $H_N^{(c)}=\sum_{k=1}^{N}\frac{1}{k^c}$ — часткова сума узагальненого гармонічного ряду. Тоді, аналогічно як у [6], одержуємо

$$E_1=\frac{1}{H_N^{(c)}}\left(H_N^{(c)}+s\,S_{ms}^{(c)}(n)-S_m^{(c)}(sn)\right),\quad E_2=\frac{1}{H_N^{(c)}}\left[(n+1)H_N^{(c)}-S_{ms}^{(c)}(n)\right],$$

де $S_{ms}^{(c)}(n)=\sum_{k=1}^{n}H_{kms}^{(c)}$, $S_m^{(c)}(n)=\sum_{k=1}^{n}H_{km}^{(c)}$.

Тому

$$E_t=\frac{1}{H_N^{(c)}}\left\{\left[(n+1)H_N^{(c)}-S_{ms}^{(c)}(n)\right]\left(a_0+t_0\right)+\left[H_N^{(c)}+s\,S_{ms}^{(c)}(n)-S_m^{(c)}(ns)\right]t_0\right\}.$$

Використовуючи для $S_{ms}^{(c)}(n)$ і $S_{m}^{(c)}(ns)$, відповідно, апроксимації [9]

$$\overline{S}_{ms}^{(c)}(n)=nH_N^{(c)}+\frac{N^{1-c}}{1-c}\left(\frac{c-1}{2-c}n+\frac{\alpha^{(c)}(n)}{n^{1-c}}\right),$$

$$\overline{S}_{m}^{(c)}(ns)=nsH_N^{(c)}+\frac{N^{1-c}}{1-c}\left(\frac{c-1}{2-c}ns+\frac{\alpha^{(c)}(ns)}{(ns)^{1-c}}\right),$$

де $\alpha^{(c)}(n)=H_n^{(c-1)}-\frac{1}{2-c}n^{2-c}$, $\alpha^{(c)}(ns)=H_{ns}^{(c-1)}-\frac{1}{2-c}(ns)^{2-c}$ — повільно зростаючі функції, з достатньо високою точністю можемо прийняти

$$E_t=\frac{1}{H_N^{(c)}}\left\{\left[H_N^{(c)}-\frac{N^{1-c}}{1-c}\left(\frac{c-1}{2-c}n+\frac{\alpha^{(c)}(n)}{n^{1-c}}\right)\right](a_0+t_0)+\right.$$
$$\left.+\left[H_N^{(c)}+\frac{N^{1-c}}{1-c}\left(s\frac{\alpha^{(c)}(n)}{n^{1-c}}-\frac{\alpha^{(c)}(ns)}{(ns)^{1-c}}\right)\right]t_0\right\}$$

або

$$E_t=\frac{1}{H_N^{(c)}}\left\{\left[H_N^{(c)}-\frac{N^{1-c}}{1-c}\left(\frac{c-1}{2-c}n+\frac{\alpha^{(c)}(n)}{n^{1-c}}\right)\right]\left(b_0+t_0+d_0\frac{N}{n}\right)+\right.$$
$$\left.+\left[H_N^{(c)}+\frac{N^{1-c}}{1-c}\left(\frac{N}{m}\,\frac{\alpha^{(c)}(n)}{n^{2-c}}-\frac{\alpha^{(c)}(N/m)}{(N/m)^{1-c}}\right)\right]t_0\right\}.$$

Візьмемо похідну від E_t по n, покладаючи $\frac{d\alpha^{(c)}(n)}{dn}\approx\alpha^{(c)}(n+1)-\alpha^{(c)}(n)$. Одержимо

$$\frac{dE_t}{dn}\approx\frac{1}{H_N^{(c)}}\left\{-\frac{N^{1-c}}{1-c}\left[\frac{c-1}{2-c}+\frac{n\left(\alpha^{(c)}(n+1)-\alpha^{(c)}(n)\right)-(1-c)\alpha^{(c)}(n)}{n^{2-c}}\right]\times\right.$$
$$\times\left(b_0+t_0+d_0\frac{N}{n}\right)+d_0\frac{N}{n^2}\left[H_N^{(c)}-\frac{N^{1-c}}{1-c}\left(\frac{c-1}{2-c}n+\frac{\alpha^{(c)}(n)}{n^{1-c}}\right)+\right.$$
$$\left.\left.+\frac{N^{1-c}}{1-c}\,\frac{N}{m}\,\frac{n\left(\alpha^{(c)}(n+1)-\alpha^{(c)}(n)\right)-(2-c)\alpha^{(c)}(n)}{n^{3-c}}t_0\right]\right\}.$$

Тоді для наближеного обчислення значення параметра n, за якого E_t приймає найменше значення, дістаємо рівняння

$$n^{3-c}\frac{b_0+t_0}{d_0}+n^{2-c}N+(2-c)N^c n^{1-c}\left[H_n^{(c)}-\frac{N^{1-c}}{1-c}\left(\frac{c-1}{2-c}n+\frac{\alpha^{(c)}(n)}{n^{1-c}}\right)\right]=$$

$$=\frac{2-c}{1-c}\left\{n\left[\alpha^{(c)}(n+1)-\alpha^{(c)}(n)\right]-(1-c)\alpha^{(c)}(n)\right\}\left(\frac{N}{m}\frac{t_0}{d_0}-n\frac{b_0+t_0}{d_0}-N\right).$$

1.5. Порівняння результатів. Оскільки на практиці в більшості випадків визначити значення сталих b_0, d_0 та t_0 доволі складно, та й значення цих величин відрізняються для різних обчислювальних машин, то приймемо надалі, що нам відомі значення відношень b_0/d_0 та t_0/d_0, які є достатньо близькі для різних обчислювальних машин. Тому замість E_t ми дослідимо функцію E_t/d_0, яка у випадку різних законів розподілу ймовірностей звертання до записів буде мати такий вигляд:

- рівномірний розподіл

$$\frac{E_t}{d_0}=\frac{1}{2}\left[\left(\frac{N}{ms}+1\right)\left(\frac{b_0}{d_0}+ms\right)+\left(\frac{N}{ms}+s+2\right)\frac{t_0}{d_0}\right];$$

- «бінарний» розподіл

$$\frac{E_t}{d_0}=\frac{2^{ms}}{2^{ms}-1}\left(\frac{b_0}{d_0}+ms\right)+\left(\frac{2^m}{2^m-1}+\frac{2^{ms}-s}{2^{ms}-1}\right)\frac{t_0}{d_0};$$

- закон Зіпфа

$$\frac{E_t}{d_0}=\frac{1}{H_N}\left\{\left(H_N+n-\frac{1}{2}\ln n-C_1\right)\left(\frac{b_0+t_0}{d_0}+\frac{N}{n}\right)+\right.$$

$$\left.+\left[H_N+\left(\frac{N}{mn}-1\right)\left(\frac{1}{2}\ln n+C_1\right)-\frac{1}{2}\ln\frac{N}{nm}\right]\frac{t_0}{d_0}\right\};$$

- узагальнений закон розподілу

$$\frac{E_t}{d_0}=\frac{1}{H_N^{(c)}}\left\{\left[H_N^{(c)}-\frac{N^{1-c}}{1-c}\left(\frac{c-1}{2-c}n+\frac{\alpha^{(c)}(n)}{n^{1-c}}\right)\right]\left(\frac{b_0+t_0}{d_0}+\frac{N}{n}\right)+\right.$$

$$\left.+\left[H_N^{(c)}+\frac{N^{1-c}}{1-c}\left(\frac{N}{m}\frac{\alpha^{(c)}(n)}{n^{2-c}}-\frac{\alpha^{(c)}(N/m)}{(N/m)^{1-c}}\right)\right]\frac{t_0}{d_0}\right\}.$$

На рис. 1 зображено залежність оптимального значення параметра n від зміни закону розподілу ймовірностей звертання до записів для $m=10$ і $N=10^6$, а на рис. 2 зображено залежність функції E_t/d_0 від зміни закону ймовірностей звертання до записів за знайдених оптимальних значеннях параметра n, $m=10$ і $N=10^6$.

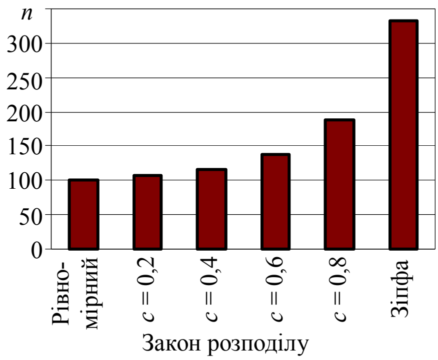

Рис. 1. Оптимальні значення параметра n за різних законів розподілу ймовірностей звертання до записів, $m = 10$ та $N = 10^6$

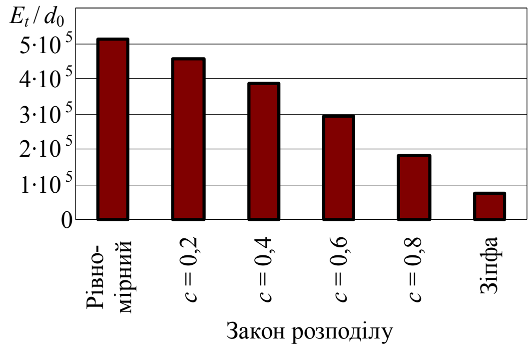

Рис. 2. Значення функції E_t / d_0 за різних законів розподілу ймовірностей звертання до записів, $m = 10$ і знайдених оптимальних значень n у випадку $N = 10^6$

Висновки. Розглянуто використання методу m-паралельного блочного пошуку для відшукання записів у послідовних упорядкованих файлах баз даних. Побудовані оптимальні стратегії пошуку записів із використанням методу m-паралельного блочного пошуку в послідовних файлах, які зберігаються у зовнішній пам'яті багатопроцесорної ЕОМ, для таких законів розподілу ймовірностей звертання до записів, як: рівномірний, «бінарний», Зіпфа й узагальнений, частковим випадком якого є розподіл, що наближено задовольняє правило «80–20». За критерій оптимальності взято математичне сподівання загального часу, необхідного для пошуку запису у файлі.

На основі одержаних даних приходимо до висновку, що оптимальні стратегії пошуку записів із використанням розглянутого варіанта методу m-паралельного блочного пошуку доволі суттєво залежать від закону розподілу ймовірностей звертання до записів.

Література

[1] *Лісовець, В.* Метод m-паралельного послідовного перегляду записів та його використання для пошуку інформації у послідовних файлах баз даних / *В. Лісовець, Г. Цегелик* // Фіз.-мат. моделювання та інформ. технології. — 2007. — Вип. 5. — С. 109-119.

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

[3] *Лісовець, В.* Метод m-паралельного послідовного пошуку записів у файлах баз даних і його ефективність / *В. Лісовець, Г. Цегелик* // Вісн. Львів. ун-ту. Сер. прикл. матем. та інформ. — 2006. — Вип. 13. — С. 177-186.

[4] *Лісовець, В.* Моделювання та оптимізація паралельного пошуку інформації у файлах баз даних / *В. Лісовець, Г. Цегелик* // Тези третьої міжнародної науково-технічна конференції: «Комп'ютерні науки та інформаційні технології». — Львів. — 2008. — С. 277-280.

[5] *Лісовець, В.* Один з варіантів методу m-паралельного блочного пошуку записів і його ефективність / *В. Лісовець, Г. Цегелик* // Фіз.-мат. моделювання та інформ. технології. — 2008. — Вип. 7. — С. 103-111.

[6] *Цегелик, Г. Г.* Системы распределенных баз данных / *Г. Г. Цегелик*. — Львов: Світ, 1990. — 168 с.
[7] *Кнут Д.* Искусство программирования для ЭВМ. Т. 3. Сортировка и поиск / *Д. Кнут*. — Москва: Изд. дом «Вильямс», 2000. — 832 с.
[8] *Мартин, Дж.* Организация баз данных в вычислительных системах / *Дж. Мартин*. — Москва: Мир, 1980. — 644 с.
[9] *Цегелик, Г. Г.* Организация и поиск информации в базах данных / *Г. Г. Цегелик*. — Львов: Вища школа, 1987. — 176 с.

Construction of the optimal strategies of information searching in sequential files of database in the case of *m*-parallel block record browsing method

Hryhoriy Tsehelyk, Volodymyr Lisovets

The using of m-parallel block record browsing method is considered for field searching in ordered files of database which are stored in external memory of multiprocessors computers. The optimal search strategies are built with using of the method of m-parallel block search for probability distribution of record request frequency as: discrete uniform, binomial, Zipf and generalized the partial case of witch is the probability distribution approximately satisfying the rule «80 – 20». The mathematical expectation of total time needed for search of a record in file is taken as a criterion of optimality. The optimal search strategies in ordered files with using of considered variant of m-parallel block record browsing method depends essentially on the law of probability distribution.

Построение оптимальных стратегий выбора информации в последовательных файлах баз данных с использованием метода *m*-параллельного блочного поиска

Григорий Цегелик, Владимир Лисовец

Рассматривается использование метода m-параллельного блочного поиска для поиска записей в упорядоченных файлах баз данных, хранящихся во внешней памяти многопроцессорной ЭВМ. При использовании m-параллельного блочного поиска строятся оптимальные стратегии поиска для таких законов распределения вероятностей обращения к записям, как: равномерный, «бинарный», Зипфа и обобщенный, частным случаем которого является распределение, которое приближенно удовлетворяет правило «80 – 20». В качестве критерия оптимальности принято математическое ожидание общего времени, необходимого для поиска записи в файле. Оптимальные стратегии поиска записей в упорядоченных файлах при использовании рассматриваемого варианта метода m-параллельного блочного поиска достаточно существенно зависят от закона распределения вероятностей обращения к записям.

Представлено професором Я. Савулою

Отримано 09.09.09