роду за допомогою квантових нечітких множин другого роду в $q_{II} f$ -системах, що складає основу успішного проведення практичних робіт.

Література

1. Пастух О.А. Квантові нечіткі множини з комплексно значною характеристичною функцією і їх використання для квантового комп'ютера / О.А.Пастух // Вісник Хмельницького національного у-ту. – 2006. – Т.1. – № 2. – С.158-161.

2. Пастух О.А. Квантова нечітка випадкова подія та її маргінальна амплітуда ймовірності / О.А.Пастух // Вісник Хмельницького національного у-ту. – 2006. – № 5. – С.58-60.

3. Пастух О.А. Повний біунарний уноїд квантових нечітких булевих підмножин на просторі $[0;\infty)$ / О.А.Пастух // Вісник Хмельницького національного у-ту. – 2007. – № 1. – С.196-198.

4. Пастух О.А. Основи зв'язку між математичними формалізмами інформаційних систем, нечітких інформаційних систем та квантових інформаційних систем / О.А.Пастух // Вісник Хмельницького національного у-ту. – 2008. – № 3. – С.87-98.

Надійшла 17.11.2009 р.

УДК 004.272.26

Г.Г. ЦЕГЕЛИК, В.Я. ЛІСОВЕЦЬ
Львівський національний університет ім. І. Франка

ПОБУДОВА ОПТИМАЛЬНИХ СТРАТЕГІЙ ПОШУКУ ІНФОРМАЦІЇ У ПОСЛІДОВНИХ ФАЙЛАХ БАЗ ДАНИХ У ВИПАДКУ ВИКОРИСТАННЯ ОДНОГО З ВАРІАНТІВ МЕТОДУ М-ПАРАЛЕЛЬНОГО БЛОЧНОГО ПОШУКУ

Будуються оптимальні стратегії пошук записів з використанням одного з варіантів методу m-паралельного блочного пошуку в послідовних упорядкованих файлах баз даних, які зберігаються в зовнішній пам'яті багатопроцесорної ЕОМ, для таких законів розподілу ймовірностей звертання до записів, як: рівномірний, „бінарний", Зіпфа та 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.

The optimal search strategies are built with using the variant of the method of m-parallel block search in ordered files of database for probability distribution of record request frequency as: discrete uniform, binomial, Zipf and generalized the partial occasion 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.

Ключові слова: стратегія пошуку, критерії оптимальності.

Вступ. Завдяки високій надійності та продуктивності багатопроцесорні ЕОМ широко використовуються для підтримки й організації великих баз даних (БД). При розв'язуванні 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, у значній мірі визначається ефективністю методів паралельного пошуку інформації в БД.

Дослідження ефективності методів пошуку інформації в базах даних є досить складною задачею. Зазвичай, за критерій ефективності береться середня кількість порівнянь, необхідних для пошуку запису. На практиці це теоретичне середнє досить часто відрізняється від реальної середньої кількості порівнянь. Насамперед це пов'язано з тим, що ймовірності звертання до записів у файлах баз даних підпорядковані нерівномірним законам розподілу: одні записи шукаються досить часто, інші дуже рідко. В даній роботі припускаємо, що ми можемо визначити закон розподілу, за яким розподілені записи. Це ми можемо зробити, наприклад, маючи статистичну інформацію конкретної бази даних.

В роботах [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 в послідовних файлах у випадку використання методів послідовного перегляду і блочного пошуку записів для однопроцесорних ЕОМ та різних законів розподілу ймовірностей звертання до записів, розглянуті в [6]. А в [4] побудовано оптимальні стратегії пошуку записів в послідовних файлах, які зберігаються у зовнішній пам'яті багатопроцесорної ЕОМ, використовуючи метод m-паралельного послідовного перегляду.

В даній роботі, використовуючи один з варіантів методу m-паралельного блочного пошуку записів [5], побудуємо оптимальні стратегії пошуку записів в файлах, які зберігаються у зовнішній пам'яті багатопроцесорної ЕОМ.

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

- „бінарний” розподіл,
- закон Зіпфа,
- узагальнений закон розподілу.

Побудова математичних моделей і виведення співвідношень для знаходження оптимальних параметрів. Нехай послідовний упорядкований файл, який містить N записів, знаходиться в зовнішній пам'яті багатопроцесорної ЕОМ, до складу якої входять m процесорів, що працюють паралельно і мають спільне поле пам'яті. Припустимо, що файл умовно розбитий на n блоків по m^2s записів в кожному, а кожний блок – на m підблоків по ms записів в кожному ($N = nm^2s$), і пошук потрібного запису відбувається так. Спочатку локалізуємо блок шляхом читання в основну пам'ять m записів, взятих по одному останньому запису з кожного підблоку фіксованого блоку. Тобто для локалізації підблоку використовується варіант методу m-паралельного блочного пошуку, що описаний в [5]. Після цього в локалізованому підблоці здійснюється пошук потрібного запису, використовуючи метод m-паралельного послідовного перегляду [1, 3].

Нехай $a_0 = b_0 + d_0m$ – час читання m 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 у вигляді суми математичного сподівання часу, необхідного для локалізації підблоку, часу, необхідного для читання підблоку в основну пам'ять і математичного сподівання часу, необхідного для пошуку запису в локалізованому блоці. Тоді E_t виразиться формулою

$$E_t = \sum_{k=1}^{n}(a_0 + t_0)k\sum_{l=1}^{m}\sum_{i=1}^{s}\sum_{j=1}^{m}p_{(k-1)m^2s-(l-1)ms+(i-1)m+j} + a_0ms +$$

$$+\sum_{k=1}^{n}\sum_{l=1}^{m}\sum_{i=1}^{s}it_0\sum_{j=1}^{m}p_{(k-1)m^2s-(l-1)ms+(i-1)m+j},$$

або

$$E_t = \sum_{k=1}^{n}\sum_{l=1}^{m}\sum_{i=1}^{s}\sum_{j=1}^{m}\left[(a_0 + t_0)k + it_0\right]p_{(k-1)m^2s-(l-1)ms+(i-1)m+j} + a_0ms .$$

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

$$E_t = (a_0 + t_0)E_2 + t_0E_1 + a_0ms ,$$

де

$$E_1 = \sum_{k=1}^{n}\sum_{l=1}^{m}\sum_{i=1}^{s}\sum_{j=1}^{m}ip_{(k-1)m^2s-(l-1)ms+(i-1)m+j},$$

$$E_2 = \sum_{k=1}^{n}\sum_{l=1}^{m}\sum_{i=1}^{s}\sum_{j=1}^{m}kp_{(k-1)m^2s-(l-1)ms+(i-1)m+j}.$$

1. Нехай розподіл імовірностей звертання до записів є рівномірним, тобто

$$p_i = \frac{1}{N},\ i = 1, 2, \dots, N.$$

Тоді

$$E_1 = \frac{1}{2}(s+1),\ E_2 = \frac{1}{2}(n+1)$$

і

$$E_t = \frac{1}{2}\left[(n+1)(a_0 + t_0) + (s+1)t_0\right] + a_0sm ,$$

або

$$E_t = \frac{1}{2}\left[(n+1)(b_0 + d_0m + t_0) + (s+1)t_0\right] + (b_0 + d_0m)sm .$$

Якщо врахувати, що $N = nm^2s$, то

$$E_t = \frac{1}{2}\left[(n+1)(b_0 + t_0 + d_0m) + \left(\frac{N}{nm^2} + 1\right)t_0\right] + (b_0 + d_0m)\frac{N}{nm}.$$

Знайдемо оптимальне значення параметра n, тобто значення, за яких E_t досягає мінімуму. Оскільки

$$\frac{dE_t}{dn}=\frac{1}{2}\left(b_0+t_0+d_0m-\frac{N}{n^2m^2}\right)-(b_0+d_0m)\frac{N}{n^2m^2},$$

то із співвідношення $\frac{dE_t}{dn}=0$ отримаємо

$$n=\frac{1}{m}\left(N\frac{t_0+2m(b_0+d_0m)}{b_0+t_0+d_0m}+1\right)^{1/2}.$$

Тоді оптимальне значення параметра s матиме такий вигляд

$$s=\frac{1}{m}\left(N\frac{b_0+t_0+d_0m}{t_0+2m(b_0+d_0m)}+1\right)^{1/2}.$$

Отже, за відомої кількості процесорів m та отриманих співвідношень для визначення оптимальних значень параметрів n та s, можемо розбити файл із записами на оптимальну кількість блоків (n), так щоб значення математичного сподівання загального часу, необхідного для пошуку запису у файлі, було мінімальним.

2. Якщо ймовірності звертання до записів задовольняють „бінарний” розподіл, тобто

$$p_i=\frac{1}{2^i},\ i=1,2,\ldots,N\text{-}1,\ p_N=\frac{1}{2^{N-1}},$$

то аналогічно як в [6]

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

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

$$E_t=a_0sm+\frac{2^{m^2s}}{2^{m^2s}-1}(a_0+t_0)+\left(\frac{2^m}{2^m-1}-\frac{s}{2^{ms}-1}\right)t_0,$$

або

$$E_t=(b_0+d_0m)sm+\frac{2^{m^2s}}{2^{m^2s}-1}(b_0+d_0m+t_0)+\left(\frac{2^m}{2^m-1}-\frac{s}{2^{ms}-1}\right)t_0.$$

Оскільки

$$\frac{dE_t}{ds}=(b_0+d_0m)m-\frac{2^{m^2s}m^2\ln 2}{(2^{m^2s}-1)^2}(b_0+d_0m+t_0)+\frac{2^{ms}(ms\ln 2-1)+1}{(2^{ms}-1)^2}t_0,$$

то для знаходження s, за якого E_t досягає мінімуму, одержуємо рівняння

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

3. Припустимо, що розподіл ймовірностей звертання до записів задовольняє закон Зіпфа, тобто

$$p_i=\frac{1}{iH_N},\quad i=1,2,\ldots,N,$$

де $H_N=\sum_{k=1}^{N}\frac{1}{k}$ – частинна сума гармонічного ряду. Тоді, аналогічно як в [6], одержуємо

$$E_1=\frac{1}{H_N}\left(H_N+s\cdot S_{ms}(mn)-S_m(mns)\right),\ E_2=\frac{1}{H_N}\left((n+1)H_N-S_{m^2s}(n)\right),$$

де

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

Тоді

$$E_t=a_0ms+\frac{1}{H_N}\left[\left((n+1)H_N-S_{m^2s}(n)\right)(a_0+t_0)+\left(H_N+s\cdot S_{ms}(mn)-S_m(mns)\right)t_0\right].$$

Використовуючи апроксимацію сум $S_{ms}(nm)$, $S_{m^2s}(n)$ і $S_m(nms)$ відповідно виразами [9]

$$\bar{S}_{ms}(nm) = nm(H_N - 1) + \frac{1}{2}\ln(nm) + C_1,$$

$$\bar{S}_{m^2 s}(n) = n(H_N - 1) + \frac{1}{2}\ln n + C_1,$$

$$\bar{S}_m(nms) = nms(H_N - 1) + \frac{1}{2}\ln(nms) + C_1,$$

де $C_1 = 0.5\ln 2\pi$, із достатньо високою точністю можемо прийняти

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

$$\left. + \left[H_N + (s-1)\left(\frac{1}{2}\ln nm + C_1\right) - \frac{1}{2}\ln s\right]t_0\right\},$$

або

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

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

Оскільки

$$\frac{dE_t}{dn} = (b_0 + d_0 m)\left(-\frac{N}{mn^2}\right) + \frac{1}{H_N}\left\{\left(1 - \frac{1}{2n}\right)(b_0 + t_0 + d_0 m) + \right.$$

$$\left. + \left[-\frac{N}{m^2 n^2}\left(\frac{1}{2}\ln nm + C_1\right) + \left(\frac{N}{m^2 n} - 1\right)\frac{1}{2n} + \frac{1}{2n}\right]t_0\right\},$$

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

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

4. Нехай імовірності звертання до записів задовольняють узагальнений закон розподілу, тобто

$$p_i = \frac{1}{i^c H_N^{(c)}}, \; i = 1,2, \dots, N,$$

де c ($0<c<1$) – будь-який параметр, $H_N^{(c)} = \sum_{k=1}^{N}\frac{1}{k^c}$ – частинна сума узагальненого гармонічного ряду.

Тоді, аналогічно як в [6], одержуємо

$$E_1 = \frac{1}{H_N^{(c)}}\left(H_N^{(c)} + s\cdot S_{ms}^{(c)}(mn) - S_m^{(c)}(mns)\right), \quad E_2 = \frac{1}{H_N^{(c)}}\left((n+1)H_N^{(c)} - S_{m^2 s}^{(c)}(n)\right),$$

де

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

Тоді

$$E_t = a_0 s + \frac{1}{H_N^{(c)}}\left\{\left[(n+1)H_N^{(c)} - S_{m^2 s}^{(c)}(n)\right](a_0 + t_0) + \left[H_N^{(c)} + s\cdot S_{ms}^{(c)}(mn) - S_m^{(c)}(mns)\right]t_0\right\}.$$

Використовуючи апроксимацію $S_{ms}^{(c)}(nm)$, $S_{m^2 s}^{(c)}(n)$ і $S_m^{(c)}(nms)$ відповідно виразами [9]

$$\bar{S}_{ms}^{(c)}(nm) = nmH_N^{(c)} + \frac{N^{1-c}}{1-c}\left(\frac{c-1}{2-c}nm + \frac{\alpha^{(c)}(nm)}{(nm)^{1-c}}\right),$$

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

$$\overline{S}_m^{(c)}(nms) = nmsH_N^{(c)} + \frac{N^{1-c}}{1-c}\left(\frac{c-1}{2-c}nms + \frac{\alpha^{(c)}(nms)}{(nms)^{1-c}}\right),$$

де

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

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

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

повільно зростаючі функції, із достатньо високою точністю одержуємо

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

$$\left. + \left[H_N^{(c)} + \frac{N^{1-c}}{1-c}\left(s\frac{\alpha^{(c)}(mn)}{(mn)^{1-c}} - \frac{\alpha^{(c)}(mns)}{(mns)^{1-c}}\right)\right]t_0\right\},$$

або

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

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

Візьмемо похідну від функції E_t по n, замінюючи $\frac{d\alpha^{(c)}(n)}{dn}$ і $\frac{d\alpha^{(c)}(mn)}{dn}$ відповідно різницями $\alpha^{(c)}(n+1) - \alpha^{(c)}(n)$ і $\alpha^{(c)}(mn+1) - \alpha^{(c)}(mn)$. Одержимо

$$\frac{dE_t}{dn} \approx -\frac{N}{mn^2}(b_0 + d_0 m) +$$

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

$$\left. + \frac{N^{2-c}}{1-c}\frac{n\alpha^{(c)}(mn+1) - (n+2-c)\alpha^{(c)}(mn)}{(mn)^{3-c}}t_0\right].$$

Для наближеного обчислення значення параметра n, за якого E_t досягає мінімуму, одержуємо рівняння

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

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

Оскільки на практиці в більшості випадків визначити значення сталих b_0, d_0 та t_0 є досить складно, та й значення цих сталих досить сильно змінюється для різних обчислювальних машин, то в подальших обчисленнях будемо припускати, що нам відомі значення відношень b_0/d_0 та t_0/d_0, які є достатньо близькими для різних обчислювальних машин. Тому для практичних обчислень дослідження функції E_t ми замінимо на дослідження функції E_t/d_0, яка у випадку розглянутого підходу до побудови оптимальних стратегій матиме такий вигляд:

- рівномірний розподіл

$$\frac{E_t}{d_0} = \frac{1}{2}\left[(n+1)\left(\frac{b_0 + t_0}{d_0} + m\right) + \left(\frac{N}{nm^2} + 1\right)\frac{t_0}{d_0}\right] + \left(\frac{b_0}{d_0} + m\right)\frac{N}{nm};$$

- „бінарний” розподіл

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

- закон Зіпфа

$$\frac{E_t}{d_0}=\left(\frac{b_0}{d_0}+m\right)\frac{N}{mn}+\frac{1}{H_N}\left\{\left(H_N+n-\frac{1}{2}\ln n-C_1\right)\left(\frac{b_0+t_0}{d_0}+m\right)+\right.$$
$$\left.+\left[H_N+\left(\frac{N}{m^2n}-1\right)\left(\frac{1}{2}\ln nm+C_1\right)-\frac{1}{2}\ln\frac{N}{m^2n}\right]\frac{t_0}{d_0}\right\};$$

- узагальнений закон розподілу

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

Оптимальні значення параметра n, за яких функції E_t/d_0 досягає мінімуму, і значення функції E_t/d_0 при знайдених оптимальних n для $N = 10^6$, $b_0/d_0 = 100$, $t_0/d_0 = 0.003$, деяких m і різних законів розподілу ймовірностей звертання до записів приведені відповідно в табл. 1 і табл. 2.

Таблиця 1

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

m	Рівномірний	Узагальнений				Зіпфа	„Бінарний”
		c=0.2	c=0.4	c=0.6	c=0.8		
1	1414	1500	1633	1868	2380	3794	1 000 000
2	1000	1061	1155	1321	1683	2683	500 000
4	707	750	816	934	1190	1897	250 000
5	632	671	730	835	1064	1697	200 000
10	447	474	516	591	753	1200	100 000
20	316	335	365	418	532	849	50 000
40	224	237	258	295	376	600	25 000
50	200	212	231	264	337	537	20 000
100	141	150	163	187	238	380	10 000

Таблиця 2

Оптимальні значення функції E_t/d_0 для різних законів розподілу ймовірностей звертання до записів та різної кількості процесорів

m	Рівномірний	Узагальнений				Зіпфа	„Бінарний”
		c=0.2	c=0.4	c=0.6	c=0.8		
1	142 889	134 721	123 763	108 196	84 934	53 312	336
2	102 053	96 220	88 395	77 279	60 669	38 091	510
4	73 592	69 387	63 745	55 731	43 758	27 483	936
5	66 461	62 664	57 570	50 333	39 521	24 827	1 155
10	49 249	46 436	42 663	37 302	29 295	18 414	2 310
20	38 008	35 838	32 927	28 793	22 618	14 230	4 920
40	31 375	29 585	27 184	23 775	18 684	11 780	11 340
50	30 075	28 360	26 059	22 792	17 915	11 291	15 150
100	28 385	26 767	24 599	21 520	16 926	10 690	40 200

На рис. 1. показана залежність функції E_t/d_0 від закону розподілу ймовірностей звертання до записів та кількості процесорів при знайдених оптимальних n у випадку $N=10^6$, $b_0/d_0 = 100$ та $t_0/d_0 = 0.003$. Як бачимо з рисунку, функція E_t/d_0 досить суттєво залежить як від закону розподілу, так і від кількості процесорів. Отже, математичне сподівання загального часу, необхідного для пошуку запису у файлі, E_t суттєво залежить і від закону розподілу ймовірностей звертання до записів, і від кількості процесорів.

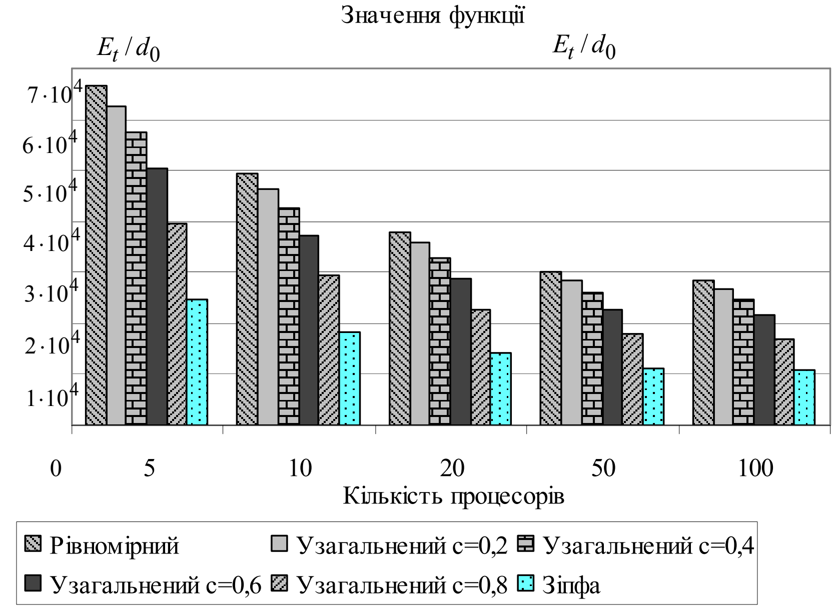

Рис. 1. Залежність функції E_t/d_0 від закону розподілу ймовірностей звертання до записів та різної кількості процесорів при знайдених оптимальних n у випадку N=10^6, b_0/d_0 = 100 та t_0/d_0 =0.003

Висновки

Розглянуто використання одного з варіантів методу m-паралельного блочного пошуку для пошуку записів у послідовних упорядкованих файлах баз даних. Побудовано оптимальні стратегії пошуку записів з 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 та 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 з використанням розглянутого варіанту метода m-паралельного блочного пошуку досить суттєво залежать як від закону розподілу ймовірностей звертання до записів, так і від кількості процесорів.

Література

1. 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.

2. 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

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

4. Volodymyr Lisovets, Hryhoriy Tsehelyk. Modeling and optimization of parallel information searching in files // International journal of computing. – 2009. – Volume 8, Issue 2. – P. 24-30.

5. Лісовець В. Я., Цегелик Г. Г. Один з варіантів методу m-паралельного блочного пошуку записів і його ефективність //Фізико-математичне моделювання та інформаційні технології. – 2008. – Вип. 7. – С. 103-111.

6. Цегелик Г.Г. Системы распределенных баз данных. – Львов: Світ, 1990, – 168 с.

7. Кнут Д. Искусство программирования для ЭВМ. Т. 3: Сортировка и поиск. – М.: Изд. Дом «Вильямс», 2000. – 832 с.

8. Мартин Дж. Организация баз данных в вычислительных системах. – М: Мир, 1980. – 644 с.

9. Цегелик Г. Г. Организация и поиск информации в базах данных. – Львов: Вища шк., 1987. – 176 с.

Надійшла 6.11.2009 р.