УДК 519.68

Х.С. Дороцька, Г.Г. Цегелик

ДО ПОБУДОВИ ОПТИМАЛЬНИХ МОДЕЛЕЙ ОДНОРІВНЕВИХ ІНДЕКСНО-ПОСЛІДОВНИХ ФАЙЛІВ

Розглядаємо однорівневий індексно-послідовний файл, який міститься в зовнішній пам'яті ЕОМ. Якщо файл містить велику кількість записів, то індекс може мати великі розміри. В цьому випадку одним з можливих шляхів прискорення пошуку запису в файлі є використання в індексі методу блочного пошуку.

Нехай N - число записів файла; m - розмір блоків записів файла; n - розмір індексу, $n=s\cdot l$, де s – кількість блоків індексу, l- кількість елементів у блоці; t_0 і t_1 - час перегляду відповідно запису файла і елемента індексу; $a_0=b_0+d_0m$ і $a_1=b_1+d_1n$ - час читання відповідно блока записів файла і індексу в основну пам'ять, де b_0, b_1, d_0, d_1 - деякі константи; p_i - ймовірність звертання до i-го запису файла; E - математичне сподівання загального часу, необхідного для пошуку запису в файлі. Тоді при використанні для пошуку в індексі методу блочного пошуку, а в файлі – послідовного перегляду для E матимемо формулу

$$E=\sum_{k=1}^{s}\sum_{i=1}^{l}\sum_{j=1}^{m}\left(a_0+a_1+(k+i)t_1+jt_0\right)p_{(k-1)ml+(i-1)m+j}\ .$$

Явний вираз для E у випадку різних законів розподілу ймовірностей звертання до записів і співвідношення для знаходження значень параметрів, при яких математичне сподівання досягає мінімуму, наведені в [1-3]. В даній статті досліджується поведінка математичного сподівання в околі точок мінімуму і залежність математичного сподівання від зміни закону розподілу ймовірностей. Наведемо окремі результати таких досліджень.

Якщо розподіл ймовірностей звертання до записів є рівномірним, то

$$E=a_0+a_1+\frac{1}{2}\left((s+l+2)t_1+(m+1)t_0\right).$$

Поведінка функції E/d_0 в околі точки мінімуму математичного сподівання показана на рис. 1.

При "бінарному" законі розподілу ймовірностей

$$E = a_0 + a_1 + \frac{l}{2^N} t_1 + \frac{m}{2^N} t_0 + \left(1 - 2^{-N}\right)\left(\left(\frac{2^{ml} - l}{2^{ml} - 1} + \frac{2^m}{2^m - 1}\right)t_1 + \left(2 - \frac{m}{2^m - 1}\right)t_0\right).$$

Поведінка функції E/d_0 в околі точки мінімуму показана на рис. 2.

Якщо ймовірності звертання до записів розподілені за законом Зіпфа, то

$$E = a_0 + a_1 + \frac{1}{H_N}\left(\left(2H_N + s + \left(\frac{1}{2}\ln s + C_1\right)(l-2) - \frac{1}{2}\ln l\right)t_1 + \frac{Nt_0}{sl}\left(\frac{1}{2}\ln sl + C_1\right)\right)$$

Поведінка функції E/d_0 в околі точки мінімуму приведена рис. 3.

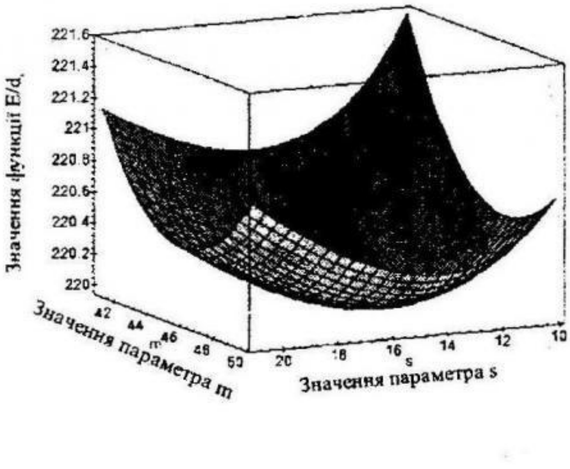


Рис. 1. Поведінка функції E/d_0 в околі точки мінімуму математичного сподівання при рівномірному розподілі ймовірностей звертання до записів (N=10000, $(b_1+b_0)/d_0$=100, t_0/d_0=0.5, t_1/d_1=2)

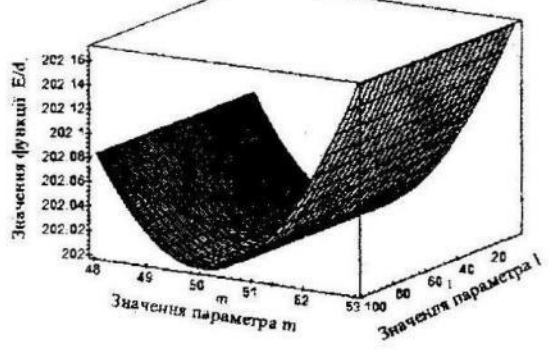


Рис. 2. Поведінка функції E/d_0 в околі точки мінімуму математичного сподівання пр "бінарному" розподілі ймовірностей звертання до записів (N=10000, $(b_1+b_0)/d_0$=100, t_0/d_0=0.5, t_1/d_1=2)

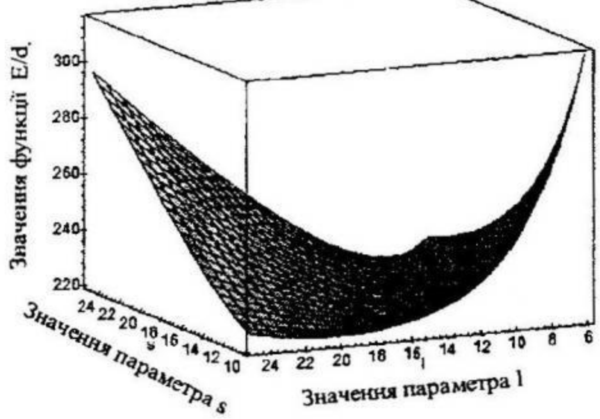
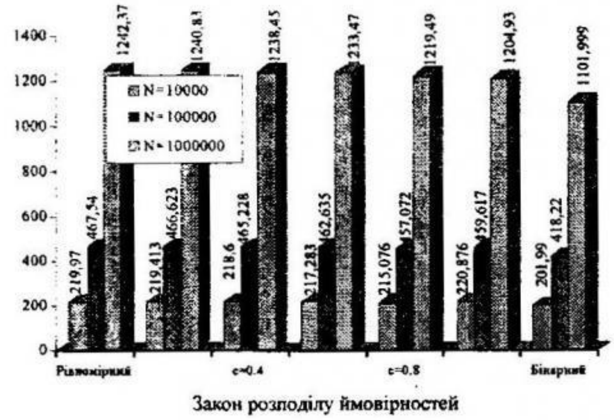


Рис. 3. Поведінка функції E/d_0 в околі точки мінімуму математичного сподівання при розподілі ймовірностей звертання до записів за законом Зіпфа (N=10000, $(b_1+b_0)/d_0$=100, t_0/d_0=0.5, t_1/d_1=2)

На діаграмі представлена залежність оптимальних значень функції E/d_0 від зміни закону розподілу ймовірностей звертання до записів



Діаграма. Залежність оптимального значення математичного сподівання від зміни закону розподілу ймовірностей звертання до записів (N=10000, t_0/d_0=0.5, t_1/d_1=2, $(b_1+b_0)/d_0$=100)

1. Ц е г е л и к Г. Г. Системы распределённых баз данных.-Львов: Свит, 1990.-168с. 2. Ц е г е л и к Г. Г. Организация и поиск информации в базах данных.-Львов: Вища шк.,1987.-176с. 3. Ц е г е л и к Г. Г. Оптимальные по времени поиска модели индексно-последовательных файлов при неравномерном распределении вероятностей обращения к записям //Программирование. 1988. №2. С.81-86.