Розв'язано задачу визначення параметрів оптимальної організації однорівневих індексопослідовних файлів баз даних, що зберігаються в зовнішній пам'яті багатопроцесорних ЕОМ, при використанні паралельного пошуку для різних законів розподілу ймовірностей звертання до записів

Ключові слова: бази даних, індексопослідовні файли, оптимізація, багатопроцесорні системи

Решена задача определения параметров оптимальной организации одноуровневых индексопоследовательных файлов баз данных, хранящихся во внешней памяти многопроцессорных ЭВМ, в случае использования параллельного поиска для разных законов распределения вероятностей обращения к записям

Ключевые слова: базы данных, индексопоследовательные файлы, оптимизация, многопроцессорные системы

Solved the problem of defining the parameters of the optimal index-sequential database files organization stored in external memory of multiprocessor computers using parallel search for various laws of probability distribution

Keywords: databases, index-sequential files, optimization, multiprocessor systems

УДК 519.68

МОДЕЛІ ОДНОРІВНЕВИХ ІНДЕКСО-ПОСЛІДОВНИХ ФАЙЛІВ БАЗ ДАНИХ ПРИ ВИКОРИСТАННІ БАГАТОПРОЦЕСОРНИХ СИСТЕМ

Г. Г. Цегелик
Доктор фізико-математичних наук, професор, завідуючий кафедрою*

Р. О. Обухівський
Аспірант*
*Кафедра математичного моделювання соціально-економічних процесів
Львівський національний університет імені Івана Франка
вул. Університетська, 1, м. Львів, 79000
Контактний тел.: 093-706-12-23
E-mail: jarlacs@gmail.com

Найбільш поширеною організацією файлів баз даних є індексопослідовна [1]. У випадку однопроцесорної ЕОМ ефективність однорівневих і багаторівневих індексопослідовних файлів для різних законів розподілу ймовірностей звертання до записів досліджена в [2-6]. В роботі дослідимо ефективність однорівневої індексопослідовної організації файлів баз даних, що зберігаються в зовнішній пам'яті багатопроцесорної ЕОМ, у випадку використання методу m-паралельного послідовного перегляду для пошуку запису і елемента індексу, відповідно, в блоці записів і індексі. За критерій ефективності візьмемо математичне сподівання загального часу, необхідного для пошуку запису у файлі. Для різних законів розподілу ймовірностей звертання до записів знайдемо явний вираз математичного сподівання і визначимо значення параметрів (розмір блоків записів файла і індекса), які мінімізуть його.

Формулювання задачі

Розглянемо однорівневий індексопослідовний файл, який знаходиться в зовнішній пам'яті ЕОМ, до складу якої входить m процесорів, що паралельно працюють і мають спільне поле пам'яті. Припустимо, що для пошуку елемента в індексі і запису у файлі використовується метод m-паралельного послідовного перегляду [6]. Нехай N – кількість записів файла; $r = mn$ – розмір індекса; $q = ml$ – розмір блоків записів файла; $a_0 = b_0 + d_0 ml$ – час читання блока записів в основну пам'ять, $a_1 = b_1 + d_1 mn$ – час читання індекса в основну пам'ять, де b_0, b_1, d_0, d_1 – деякі сталі; t_0 – час виконання операції m-паралельного перегляду записів в основній пам'яті, t_1 – час виконання операції m-паралельного перегляду елементів індекса в основній пам'яті; p_i – ймовірність звертання до i-го запису файла; E – математичне сподівання загального часу, необхідного для пошуку запису у файлі. Тоді

$$E = \sum_{s=1}^{n}\sum_{k=1}^{m}\sum_{i=1}^{l}\sum_{j=1}^{m}\left(a_0 + a_1 + st_1 + it_0\right)p_{\varphi(s,k,i,j)},$$

де

$$\varphi(s,k,i,j) = (s-1)m^2 l + (k-1)ml + (i-1)m + j.$$

Знайдемо явний вираз для E у випадку різних законів розподілу ймовірностей звертання до записів і визначимо значення параметрів n і l, за яких E досягає мінімуму.

1. Рівномірний розподіл імовірностей звертання до записів

Якщо розподіл імовірностей звертання до записів є рівномірним, то для E одержуємо вираз

$$E = a_0 + a_1 + \frac{1}{2}\left((n+1)t_1 + (l+1)t_0\right),$$

або

$$E = b_0 + \frac{d_0 N}{nm} + b_1 + d_1 nm + \frac{1}{2}\left((n+1)t_1 + \left(\frac{N}{m^2 n} + 1\right)t_0\right).$$

Оскільки

$$\frac{dE}{dn} = -\frac{d_0 N}{n^2 m} + d_1 m + \frac{1}{2}\left(t_1 - \frac{N t_0}{n^2 m^2}\right),$$

то функція E досягає мінімуму при

$$n = \frac{1}{m}\left(\frac{(2d_0 m + t_0)N}{2d_1 m + t_1}\right)^{\frac{1}{2}}, l = \frac{1}{m}\left(\frac{(2d_1 m + t_1)N}{2d_0 m + t_0}\right)^{\frac{1}{2}}.$$

Оскільки на практиці в більшості випадків визначити значення сталих d_0, d_1, t_0 і t_1 є досить складно, до того ж значення цих сталих сильно відрізняються для різних обчислювальних машин, то в подальших обчисленнях будемо припускати, що нам відомі значення відношень $\frac{t_0}{d_0}$, $\frac{t_1}{d_0}$, $\frac{d_1}{d_0}$ і $\frac{b_0 + b_1}{d_0}$, які є достатньо близькими для різних машин. Тому дослідження функції E ми замінимо на дослідження функції $\frac{E}{d_0}$.

В табл. 1 наведені значення l, n і E/d_0 для різних значень d_1/d_0, t_0/d_0 та m у випадку рівномірного розподілу, за яких математичне сподівання досягає мінімуму.

Оптимальні значення параметрів l і n, а також E/d_0 в залежності від t_0/d_0, d_1/d_0 та m у випадку $t_0 = t_1$, $\frac{b_0 + b_1}{d_0} = 1000$ і $N = 10^6$.

2. "Бінарний" розподіл імовірностей звертань до записів

Нехай імовірності звертання до записів задовольняють "бінарний" розподіл [7]. Тоді, аналогічно як в [5], для E одержуємо вираз

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

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

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

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

або

$$E = b_0 + d_0 ml + b_1 + \frac{d_1 N}{ml} + \frac{2^{m^2 l}}{2^{m^2 l} - 1}t_1 + \left(\frac{2^m}{2^m - 1} - \frac{l}{2^{ml} - 1}\right)t_0.$$

Оскільки

$$\frac{dE}{dl} = d_0 m - \frac{d_1 N}{ml^2} - \frac{2^{m^2 l} m^2 \ln 2}{\left(2^{m^2 l} - 1\right)^2}t_1 + \frac{2^{ml}(ml\ln 2 - 1) + 1}{\left(2^{ml} - 1\right)^2}t_0,$$

то для визначення значення параметра l, за якого E досягає мінімуму, дістаємо рівняння

$$\frac{2^{ml}}{\left(2^{ml} - 1\right)^2}\left(ml\ln 2 - 1 + 2^{-ml}\right)\frac{t_0}{t_1} -$$

$$-\frac{2^{m^2 l}}{\left(2^{m^2 l} - 1\right)^2}m^2 \ln 2 = \frac{d_1 N}{t_1 ml^2} - \frac{d_0 m}{t_1}.$$

В табл. 2 наведені значення l, n і E/d_0 для різних значень d_1/d_0, t_0/d_0 та m у випадку "бінарного" розподілу, за яких математичне сподівання досягає мінімуму.

Таблиця 1

$\frac{d_1}{d_0}$	m	t_0/d_0								
		10			100			1000		
		l	n	E/d_0	l	n	E/d_0	l	n	E/d_0
10	8	319.6	48.9	9320.4	187.1	83.5	22808.3	133.6	117.0	137701.9
	16	175.2	22.3	8368.0	111.5	35.0	15816.1	70.7	55.3	74946.9
	32	92.6	10.5	7863.7	66.4	14.7	11986.5	38.8	25.2	43280.3
100	8	983.6	15.9	26584.6	478.5	32.7	56609.0	200.0	78.1	205162.5
	16	546.4	7.1	23958.7	312.5	12.5	42350.0	126.1	31.0	132120.1
	32	290.8	3.4	22532.6	196.7	5.0	33364.8	82.4	11.8	89687.3
1000	8	3101.8	5.0	81657.8	1472.6	10.6	171925.2	511.3	30.6	521494.9
	16	1725.4	2.3	73478.2	974.6	4.0	129752.9	353.4	11.1	366734.5
	32	919.1	1.1	69022.7	617.8	1.6	102421.3	244.3	4.0	261882.8

Таблиця 2

$\frac{d_1}{d_0}$	m	t_0/d_0								
		10			100			1000		
		l	n	E/d_0	l	n	E/d_0	l	n	E/d_0
10	8	395.3	39.5	7344.6	395.4	39.5	7524.9	396.9	39.4	9328.5
	16	197.6	19.8	7344.6	197.7	19.8	7524.6	197.6	19.8	9324.6
	32	98.8	9.9	7344.6	98.8	9.9	7524.6	98.9	9.9	9324.6
100	8	1249.9	12.5	21020.0	1249.5	12.5	21200.4	1251.4	12.5	23003.9
	16	625.0	6.3	21020.0	624.8	6.3	21200.0	625.8	6.2	23000.0
	32	312.5	3.1	21020.0	312.5	3.1	21200.0	313.0	3.1	23000.0
1000	8	3952.7	4.0	64265.6	3952.2	4.0	64445.9	3967.4	3.9	66249.9
	16	1976.4	2.0	64265.6	1976.2	2.0	64445.6	1976.2	2.0	66245.6
	32	988.2	1.0	64265.6	988.2	1.0	64445.6	988.2	1.0	66245.6

Оптимальні значення параметрів l і n, а також E/d_0 в залежності від t_0/d_0, d_1/d_0 та m у випадку $t_0=t_1$, $\frac{b_0+b_1}{d_0}=1000$ і $N=10^6$.

3. Розподіл імовірностей звертань до записів за законом Зіпфа

Припустимо, що ймовірності звертання до записів задовольняють закон Зіпфа [7]. Тоді, аналогічно як в [5], для E одержуємо вираз

$$E=a_0+a_1+\frac{1}{H_N}\Big(\big((n+1)H_N-S_{m^2l}(n)\big)t_1+\big(lS_{ml}(mn)+H_N-S_m(mnl)\big)t_0\Big),$$

де

$$S_{m^2l}(n)=\sum_{k=1}^{n}H_{km^2l},\quad S_{ml}(mn)=\sum_{k=1}^{mn}H_{kml},\quad S_m(mnl)=\sum_{k=1}^{mnl}H_{km}.$$

Використовуючи апроксимацію $S_{m^2l}(n)$, $S_{ml}(mn)$, $S_m(mnl)$ відповідно виразами [4]

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

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

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

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

$$E=a_0+a_1+\frac{1}{H_N}\Big((H_N+n-0{,}5\ln n-C_1)t_1+\big(H_N+(l-1)(0{,}5\ln mn+C_1)-0{,}5\ln l\big)t_0\Big),$$

або

$$E=b_0+\frac{d_0N}{mn}+b_1+d_1mn+\frac{1}{H_N}\Bigg((H_N+n-0{,}5\ln n-C_1)t_1+\left(H_N+\left(\frac{N}{m^2n}-1\right)(0{,}5\ln mn+C_1)-0{,}5\ln\frac{N}{m^2n}\right)t_0\Bigg).$$

Оскільки

$$\frac{dE}{dn}=-\frac{d_0N}{mn^2}+d_1m+\frac{1}{H_N}\left(\left(1-\frac{1}{2n}\right)t_1+\frac{Nt_0}{2m^2n^2}(1-\ln mn-2C_1)\right),$$

то для визначення параметра n, за якого функція E досягає мінімуму, маємо рівняння

$$\left(2n\left(1+mH_N\frac{d_1}{t_1}\right)-1\right)nm^2=\left(2\frac{d_0}{t_1}H_Nm+\frac{t_0}{t_1}(\ln mn+2C_1-1)\right)N.$$

В табл. 3 наведені значення l, n і E/d_0 для різних значень d_1/d_0, t_0/d_0 та m у випадку закону розподілу Зіпфа, за яких математичне сподівання досягає мінімуму.

Оптимальні значення параметрів l і n, а також E/d_0 в залежності від t_0/d_0, d_1/d_0 та m у випадку $t_0=t_1$, $\frac{b_0+b_1}{d_0}=1000$ і $N=10^6$.

Таблиця 3

$\frac{d_1}{d_0}$	m	t_0/d_0								
		10			100			1000		
		l	n	E/d_0	l	n	E/d_0	l	n	E/d_0
10	8	349.2	44.7	8356.3	201.9	77.4	15288.7	89.9	173.9	57425.5
	16	185.1	21.1	7860.4	126.8	30.8	11858.7	55.6	70.3	36523.9
	32	95.5	10.2	7603.4	75.5	12.9	9855.0	35.7	27.3	24905.1
100	8	1122.3	13.9	23700.0	657.1	23.8	41796.4	230.1	67.9	128402.8
	16	590.8	6.6	22387.7	415.3	9.4	32810.3	160.5	24.3	89192.1
	32	303.6	3.2	21709.6	245.9	4.0	27474.0	111.3	8.8	63661.1
1000	8	3622.8	4.3	71289.6	2240.4	7.0	120622.6	766.5	20.4	358032.3
	16	1889.3	2.1	67837.3	1394.3	2.8	95688.1	546.3	7.1	251203.7
	32	965.8	1.0	66065.6	810.0	1.2	81147.4	382.0	2.6	179729.4

4. Узагальнений розподіл імовірностей звертань до записів

Нехай імовірності звертання до записів задовольняють узагальнений закон розподілу, де $0<c<1$ [7]. Тоді, аналогічно як в [5], для E одержуємо вираз

$$E=a_0+a_1+\frac{1}{H_N^{(c)}}\left(\left((n+1)H_N^{(c)}-S_{m^2l}^{(c)}n\right)t_1+\left(lS_{ml}^{(c)}(mn)+H_N^{(c)}-S_m^{(c)}(mnl)\right)t_0\right),$$

де

$$S_{m^2l}^{(c)}(n)=\sum_{k=1}^{n}H_{km^2l}^{(c)},\quad S_{ml}^{(c)}(mn)=\sum_{k=1}^{mn}H_{kml}^{(c)},$$

$$S_m^{(c)}(mnl)=\sum_{k=1}^{mnl}H_{km}^{(c)}.$$

Використовуючи апроксимацію $S_{m^2l}^{(c)}(n)$, $S_{ml}^{(c)}(mn)$, $S_m^{(c)}(mnl)$ відповідно виразами

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

$$\bar S_{ml}^{(c)}(mn)=mnH_N^{(c)}+\frac{N^{1-c}}{1-c}\left(\frac{c-1}{2-c}mn+\frac{\alpha^{(c)}(mn)}{(mn)^{1-c}}\right),$$

$$\bar S_{m}^{(c)}(mnl)=mnlH_N^{(c)}+\frac{N^{1-c}}{1-c}\left(\frac{c-1}{2-c}mnl+\frac{\alpha^{(c)}(mnl)}{(mnl)^{1-c}}\right),$$

де $\alpha^{(c)}(k)=H_k^{(c-1)}-\frac{1}{2-c}k^{2-c}$ – повільно зростаюча функція, з достатньо високою точністю можемо прийняти

$$E=a_0+a_1+\frac{1}{H_N^{(c)}}\left(\left(H_N^{(c)}-\frac{N^{1-c}}{1-c}\left(\frac{c-1}{2-c}n+\frac{\alpha^{(c)}(n)}{n^{1-c}}\right)\right)t_1+\right.$$
$$\left.+\left(H_N^{(c)}+\frac{N^{1-c}}{1-c}\left(\frac{\alpha^{(c)}(mn)}{(mn)^{1-c}}l-\frac{\alpha^{(c)}(mnl)}{(mnl)^{1-c}}\right)\right)t_0\right),$$

або

$$E=b_0+\frac{d_0N}{mn}+b_1+d_1mn+$$
$$+\frac{1}{H_N^{(c)}}\left(\left(H_N^{(c)}-\frac{N^{1-c}}{1-c}\left(\frac{c-1}{2-c}n+\frac{\alpha^{(c)}(n)}{n^{1-c}}\right)\right)t_1+\right.$$
$$\left.+\left(H_N^{(c)}+\frac{N^{1-c}}{1-c}\left(\frac{\alpha^{(c)}(mn)}{(mn)^{1-c}}\frac{N}{m^2n}-\frac{\alpha^{(c)}(N/m)}{(N/m)^{1-c}}\right)\right)t_0\right).$$

Замінюючи $\frac{d\alpha^{(c)}(n)}{dn}$ і $\frac{d\alpha^{(c)}(mn)}{dn}$ відповідно різницями $\alpha^{(c)}(n+1)-\alpha^{(c)}(n)$ і $\alpha^{(c)}(mn+1)-\alpha^{(c)}(mn)$ одержуємо

$$\frac{dE}{dn}\approx-\frac{d_0N}{mn^2}+d_1m+$$
$$+\frac{1}{H_N^{(c)}}\left(-\frac{N^{1-c}}{1-c}\left(\frac{c-1}{2-c}+\frac{n\alpha^{(c)}(n+1)-(n+1-c)\alpha^{(c)}(n)}{n^{2-c}}\right)t_1+\right.$$
$$\left.+\frac{N^{2-c}}{1-c}\frac{n\alpha^{(c)}(mn+1)-(n+2-c)\alpha^{(c)}(mn)}{(mn)^{3-c}}t_0\right).$$

Для визначення параметра n, за якого функція E досягає мінімуму, маємо рівняння

$$\left(n^{3-c}-\frac{2-c}{1-c}n\left(n\alpha^{(c)}(n+1)-(n+1-c)\alpha^{(c)}(n)\right)\right)\frac{t_1}{d_0}=$$
$$=(2-c)N^{c-1}n^{3-c}H_N^{(c)}\left(\frac{N}{mn^2}-m\frac{d_1}{d_0}\right)-$$
$$-\frac{2-c}{1-c}\frac{N}{m^{3-c}}\left(n\alpha^{(c)}(mn+1)-(n+2-c)\alpha^{(c)}(mn)\right)\frac{t_0}{d_0}.$$

В табл. 4, 5, 6, 7 наведені значення l, n і E/d_0 для різних значень c, d_1/d_0, t_0/d_0 та m у випадку узагальненого розподілу, за яких математичне сподівання досягає мінімуму.

Оптимальні значення параметрів l і n, а також E/d_0 в залежності від t_0/d_0, d_1/d_0 та m у випадку $c=0.2$, $t_0=t_1$, $\frac{b_0+b_1}{d_0}=1000$ і $N=10^6$.

Таблиця 4

d_1/d_0	m	t_0/d_0 = 10			100			1000		
		l	n	E/d_0	l	n	E/d_0	l	n	E/d_0
10	8	282.8	55.3	9357.5	144.7	108.0	22969.8	97.7	160.0	136119.1
	16	161.1	24.3	8383.9	87.8	44.5	16063.6	51.5	75.9	75527.0
	32	88.0	11.1	7868.3	54.3	18.0	12118.6	28.5	34.3	43706.9
100	8	874.1	17.9	26739.4	376.5	41.5	58070.8	150.6	103.8	208741.7
	16	504.0	7.8	24013.2	249.0	15.7	43427.9	94.6	41.3	135790.4
	32	278.9	3.5	22455.0	161.9	6.0	33955.6	62.1	15.7	93164.6
1000	8	2776.9	5.6	81707.7	1168.2	13.4	175901.4	391.8	39.9	539237.3
	16	1601.8	2.4	73339.8	781.2	5.0	132534.7	268.2	14.6	380725.4
	32	890.8	1.1	68301.8	512.2	1.9	103901.2	186.0	5.3	271241.6

Таблиця 5

$\frac{d_1}{d_0}$	m	t_0/d_0								
		10			100			1000		
		l	n	E/d_0	l	n	E/d_0	l	n	E/d_0
10	8	290.0	53.9	9282.3	148.1	105.5	22026.6	96.4	162.1	124050.6
	16	164.5	23.8	8342.6	90.8	43.0	15556.2	51.3	76.2	69274.1
	32	89.3	10.9	7850.6	56.2	17.4	11937.3	28.8	33.9	41737.4
100	8	905.7	17.3	26462.0	396.8	39.4	56780.3	154.9	100.9	198953.2
	16	516.5	7.6	23886.3	262.6	14.9	42654.7	98.7	39.6	131281.9
	32	281.5	3.5	22471.7	169.8	5.8	33555.4	65.4	14.9	90851.5
1000	8	2904.6	5.4	80313.8	1237.6	12.6	172589.8	415.3	37.6	525013.5
	16	1643.7	2.4	72855.8	833.2	4.7	129321.2	285.4	13.7	371858.5
	32	892.3	1.1	68682.4	538.7	1.8	102315.3	197.8	4.9	266361.2

Оптимальні значення параметрів l і n, а також E/d_0 в залежності від t_0/d_0, d_1/d_0 та m у випадку $c=0.4$, $t_0=t_1$, $\frac{b_0+b_1}{d_0}=1000$ і $N=10^6$.

Таблиця 6

$\frac{d_1}{d_0}$	m	t_0/d_0								
		10			100			1000		
		l	n	E/d_0	l	n	E/d_0	l	n	E/d_0
10	8	299.0	52.3	9179.8	152.3	102.6	20956.8	92.8	168.4	110765.2
	16	168.0	23.3	8289.5	94.6	41.3	15036.0	50.4	77.5	62730.3
	32	90.6	10.8	7818.8	58.7	16.6	11667.1	28.9	33.8	38319.0
100	8	939.8	16.6	26222.6	423.7	36.9	55373.5	160.3	97.5	188317.6
	16	529.6	7.4	23747.8	280.3	13.9	41655.6	104.2	37.5	125463.4
	32	286.6	3.4	22357.2	180.6	5.4	32816.2	69.8	14.0	86645.7
1000	8	3045.9	5.1	78859.7	1344.0	11.6	166698.8	446.4	35.0	508998.4
	16	1688.6	2.3	72345.1	892.8	4.4	126241.1	309.4	12.6	359404.9
	32	917.2	1.1	67841.8	578.6	1.7	99352.4	215.5	4.5	256345.1

Оптимальні значення параметрів l і n, а також E/d_0 в залежності від t_0/d_0, d_1/d_0 та m у випадку $c=0.6$, $t_0=t_1$, $\frac{b_0+b_1}{d_0}=1000$ і $N=10^6$.

Таблиця 7

$\frac{d_1}{d_0}$	m	t_0/d_0								
		10			100			1000		
		l	n	E/d_0	l	n	E/d_0	l	n	E/d_0
10	8	314.9	49.6	8963.9	161.5	96.8	19429.5	87.1	179.4	717.4
	16	174.1	22.4	8145.6	102.1	38.3	13937.0	49.5	78.9	52235.1
	32	92.5	10.6	7744.6	63.3	15.4	11074.8	29.6	33.0	32493.2
100	8	1000.0	15.6	25480.8	478.9	32.6	51777.4	172.9	90.4	171269.0
	16	553.1	7.1	23260.4	315.7	12.4	38877.1	115.7	33.8	113919.2
	32	294.7	3.3	22046.9	200.3	4.9	31094.3	78.9	12.4	79864.1
1000	8	3285.9	4.8	75531.6	1562.4	10.0	152453.7	514.4	30.4	469239.5
	16	1783.8	2.2	70110.2	1041.4	3.8	114654.7	361.3	10.8	326860.6
	32	944.5	1.0	66729.0	650.9	1.5	92876.5	252.0	3.9	233798.4

Оптимальні значення параметрів l і n, а також E/d_0 в залежності від t_0/d_0, d_1/d_0 та m у випадку $c=0.8$, $t_0=t_1$, $\frac{b_0+b_1}{d_0}=1000$ і $N=10^6$.

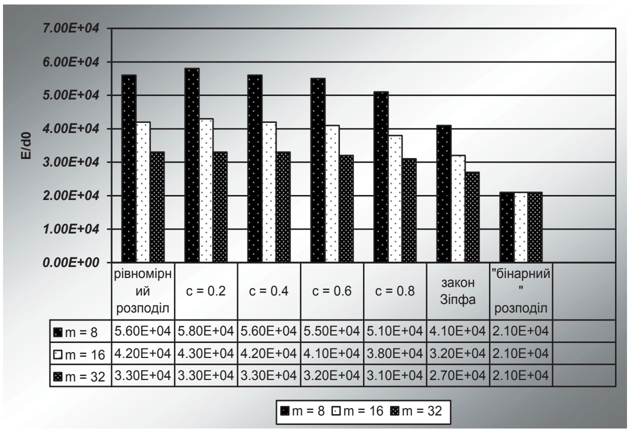

Рис. 1. Значення E/d_0 для різних законів розподілу ймовірностей для деяких m у випадку $\frac{d_1}{d_0} = \frac{t_0}{d_0} = 100$, $t_0 = t_1$, $\frac{b_0 + b_1}{d_0} = 1000$ і $N = 10^6$

Порівняння E/d_0 в залежності від закону розподілу ймовірностей звертань до записів для деяких m наведене на рис. 1. Як бачимо з рисунку, функція E/d_0, а отже і функція E, досить суттєво залежить як від закону розподілу ймовірностей звертань до записів, так і від кількості процесорів (за вийнятком "бінарного" розподілу).

Висновки

Побудовано оптимальні схеми пошуку записів в однорівневих індексопослідовних файлах, які зберігаються у зовнішній пам'яті ЕОМ, до складу якої входить m процесорів, що паралельно працюють і мають спільне поле пам'яті, для таких законів розподілу ймовірностей звертання до записів, як рівномірний, "бінарний", Зіпфа, узагальнений, частковим випадком якого є розподіл, що наближено задовільняє правило "80-20". За критерій оптимальності взято математичне сподівання загального часу, необхідного для пошуку запису у файлі.

Проведено порівняння оптимальних схем для розглянутих законів розподілу ймовірностей звертань до записів.

Література

1. Мартин Д. Организация баз данных в вычислительных системах. – М: Мир, 1980. – 662 с.
2. Цегелик Г. Г. Оптимальные по времени поиска модели индекно-последовательных файлов при неравномерном распределении вероятностей обращения к записям // Программирование. – 1988. – № 2. – С. 81–86.
3. Цегелик Г. Г. Определение параметров оптимальной организации многоуровневых индексно-последовательных файлов // Кибернетика. – 1988. – № 2. – С. 74–78.
4. Цегелик Г. Г. Организация и поиск информации в базах данных. – Львов: Вища школа, 1987. – 176 с.
5. Цегелик Г. Г. Системы распределенных баз даных. – Львов: Свит, 1990. – 168 с.
6. Цегелик Г. Г. Моделювання та оптимізація доступу до інформації файлів баз даних для однопроцесорних і багатопроцесорних систем. – Львів: Вид. ЛНУ імені Івана Франка, 2010. – 192 с.
7. Кнут Д. Искусство програмирования для ЭВМ. Т. 3. Сортировка и поиск. – М: Изд. дом. "Вильямс", 2000. – 824 с.