6. Huang W.K., Meyer F.J., and Lombardi F. An Approach for Detecting Multiple Faulty FPGA Logic Blocks, IEEE Trans. on Computers, Vol. 49, No. 1, 2000. – pp. 48-54.

7. Lakamraju V. and TessierR. Tolerating Operational Faults in Cluster-based FPGAs. Proc. ACM/SIGDA International Symp. on FPGAs, Febr. 2000. – pp. 187-194.

8. Wang, S. J., et al., Test and diagnosis of faulty logic blocks in FPGAs. Proc. ICCAD, 1997. – pp. 722-727.

9. Соловьев В.В.. Программируемые логические интегральные схемы и их применение / Соловьев В.В., Васильев А.Г. – Мн.: Беларуская навука, 1998. – 266с.

10. Локазюк В.М. Надійність, контроль, діагностика і модернізація ПК: [посібник] / Локазюк В.М., Савченко Ю.Г. – К.: Видавничий центр «Академія», 2004. – 376 с.

11. Надежность технических систем: справочник / [Ю.К. Беляев, В.А. Богатырев, В.В. Болотин и др.]; под ред. И.А. Ушаков. – М.: Радио и связь, 1985. – 608 с.

12. Altera Corporation. Reliability Report 50. – Q3 2010. – 50 p.

13. Atmel. Reliability Monitor Report. 2010. – 14 p.

14. Xilinx. Device Reliability Report. (ug116). 2010. – 106 p.

15. Erik Chmelar The Test and Diagnosis of FPGAs. A dissertation submitted to the department of electrical engineering and the committee on graduate studies of Stanford university. Stanford University. 2004. – 112 p.

16. Abramovici, M., C. Stroud, “BIST-Based Detection and Diagnosis of Multiple Faults in FPGAs, ” Proc. Int’l Test Conf., 2000. – pp. 119-141.

17. Subhasish M. Reconfigurable architecture for autonomous self-repair / M. Subhasish, W. – J. Huang, N.R. Saxena, E.J. McCluskey // IEEE Design & Test of Computers. – Volume 21, Issue 3. – 2004. – P. 228-240.

18. Хаханов В.И. Сервисное обслуживание современных цифровых систем на кристаллах / Хаханов В.И., Литвинова Е.И., Ngene Christopher Umeran // Радіоелектронні комп’ютерні системи. – 2009. – № 7. – С. 319-323.

19. Ross R. A FPGA Simulation Using Asexual Genetic Algorithms for Integrated Self-Repair / R. Rose, R. Hall // Adaptive Hardware and Systems, 2006. – AHS 2006. – First NASA/ESA Conference on Volume. – Issue 15-18 June 2006. – P. 301-304.

20. HabermannS., KotheR., Vierhaus H.T. Built-in self repair by reconfiguration of FPGAs // Proceedings of the 12th IEEE International Symposium on On-Line Testing. – 2006. – P. 187-188.

Надійшла 7.11.2010 р.

УДК 519.68

Г.Г. ЦЕГЕЛИК, Р.О. ОБУХІВСЬКИЙ
Львівський національний університет імені Івана Франка

ВИЗНАЧЕННЯ ПАРАМЕТРІВ ОПТИМАЛЬНОЇ ОРГАНІЗАЦІЇ ІНДЕКСО-ПОСЛІДОВНИХ ФАЙЛІВ БАЗ ДАНИХ

Розв’язана задача визначення параметрів (кількості рівнів індекса, розмірів блоків індекса і записів файлу) оптимальної організації індексо-послідовних файлів баз даних для різних законів розподілу ймовірностей звертання до записів. За критерій оптимальності прийнято математичне сподівання загального часу, необхідного для пошуку запису у файлі.

Solved the problem of defining the parameters (number of index levels, size of blocks and index entries of the file) of the optimal index-sequential database files for different probability distribution laws to access records. By optimality criterion adopted general expectation of time needed to find a record in the file.

Ключові слова: рівень індексу, база даних, розподіл ймовірностей.

Найбільш поширеною організацією файлів баз даних є індексо-послідовна організація [1]. При заданій кількості рівнів індекса її ефективність для різних законів розподілу ймовірностей звертання до записів досліджена в [2–6]. В роботі досліджено ефективність індексо-послідовної організації файлів залежно від трьох параметрів: кількості рівнів індекса, розміру блоків індекса і розміру блоків записів файлу. 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 визначаємо значення параметрів, які мінімізуть його.

Формулювання задачі. Розглянемо r -рівневий ($r \geq 1$) індексо-послідовний файл з однаковим розміром блоків індексу на всіх рівнях. Нехай N – кількість записів файла; l і m – розмір, відповідно, блоків індекса і блоків записів файла; $a_0 = b_0 + d_0 m$ і $a_1 = b_1 + d_1 l$ – час доступу, відповідно, до блоку записів файлу і блоку елементів індексу, де b_0, b_1, d_0, d_1 – деякі сталі; t_0 і t_1 – час перегляду, відповідно, запису файлу і елемента індекса; p_i – ймовірність звертання до i -го запису файлу. Тоді при використанні в індексі і файлі методу послідовного перегляду математичне сподівання загального часу пошуку запису виражається наступною формулою:

$$E=\sum_{i_r=1}^{l}\dots\sum_{i_1=1}^{l}\sum_{i_0=1}^{m}t(i_0,i_1,\dots,i_r)p_{\varphi(i_0,i_1,\dots,i_r)},$$

де

$$t(i_0,i_1,\dots,i_r)=a_0+a_1r+t_0i_0+t_1\sum_{k=1}^{r}i_k,$$

$$\varphi(i_0,i_1,\dots,i_r)=i_0+m\sum_{j=1}^{r}(i_j-1)l^{j-1}.$$

Знайдемо явний вираз для E у випадку різних законів розподілу ймовірностей звертання до записів і визначимо значення параметрів r, m і l, за яких математичне сподівання досягає мінімуму. Математичне сподівання E подамо у вигляді

$$E=a_0+a_1r+t_0E_0+t_1\sum_{k=1}^{r}E_k,$$

де

$$E_k=\sum_{i_r=1}^{l}\dots\sum_{i_1=1}^{l}\sum_{i_0=1}^{m}i_kp_{\varphi(i_0,i_1,\dots,i_r)},\quad k=0,1,\dots,r.$$

Рівномірний розподіл імовірностей звертання до записів

Нехай розподіл імовірностей звертання до записів є рівномірним. Тоді

$$E_0=\frac{1}{2}(m+1),\qquad E_k=\frac{1}{2}(l+1)\quad(k=1,2,\dots,r)$$

і

$$E=a_0+a_1r+\frac{1}{2}\left((m+1)t_0+r(l+1)t_1\right)$$

або

$$E=b_0+\frac{d_0N}{l^r}+r(b_1+d_1l)+\frac{1}{2}\left(\left(\frac{N}{l^r}+1\right)t_0+r(l+1)t_1\right).$$

Для визначення значень параментів l та r, за яких E досягає мінімуму, отримуємо систему рівнянь

$$\begin{cases}(2d_1+t_1)l^{r+1}=N(2d_0+t_0),\\ (2b_1+t_1+(2d_1+t_1)l)l^r=N(2d_0+t_0)\ln l\end{cases}$$

або

$$\begin{cases}\ln l=1+\dfrac{1}{l}\dfrac{2b_1+t_1}{2d_1+t_1},\\ r=\log_l\left(\dfrac{N(2d_0+t_0)}{2d_1+t_1}\right)-1.\end{cases}\tag{1}$$

Зокрема, якщо $a_0=const$, $a_1=const$, то система рівнянь для визначення значень параметрів r і l, за яких E досягає мінімуму, матиме вигляд

$$\begin{cases}\ln l=1+\dfrac{1}{l}\left(1+\dfrac{2a_1}{t_1}\right),\\ r=\log_l\left(\dfrac{Nt_0}{t_1}\right)-1.\end{cases}$$

Таблиця 1

Розв'язки системи (1) залежно від $\frac{b_1}{t_1}$ та $\frac{d_1}{t_1}$ у випадку $t_0=t_1$, $\frac{d_0}{t_0}=1000$ і $N=10^6$

$\frac{b_1}{t_1}$	$\frac{d_1}{t_1}$		
	10	100	1000
10	$l=3{,}59; r=13{,}37$	$l=2{,}82; r=14{,}54$	$l=2{,}73; r=12{,}76$
100	$l=8{,}44; r=7{,}61$	$l=3{,}59; r=11{,}60$	$l=2{,}82; r=12{,}34$
1000	$l=36{,}63; r=4{,}10$	$l=8{,}62; r=6{,}48$	$l=3{,}59; r=9{,}81$

Таблиця 2

Оптимальні значення $\frac{E}{d_0}$ залежно від $\frac{b_1}{t_1}$ та $\frac{d_1}{t_1}$ у випадку $t_0 = t_1$, $\frac{d_0}{t_0} = 1000$, $\frac{b_0}{d_0} = 1000$ і $N = 10^6$

$\frac{b_1}{t_1}$	$\frac{d_1}{t_1}$		
	10	100	1000
10	1682,76	5558,13	38708,01
100	2528,98	6715,53	39837,22
1000	7067,15	13964,65	49637,98

“Бінарний” розподіл імовірностей звертань до записів

Припустимо, що ймовірності звертань до записів відповідають “бінарному” розподілу [7], тобто:

$$p_i = \frac{1}{2^i}\ (i = 1, 2, \dots, N-1),\ p_N = \frac{1}{2^{N-1}}.$$

В цьому випадку

$$p_{\varphi(t_0, t_1, \dots, t_r)} = 2^{-\varphi(t_0, t_1, \dots, t_r)},$$

для $t_0 + t_1 + \dots + t_r < m + rl,\ 1 \le t_j \le l\ (j = 1, 2, \dots, r),\ 1 \le t_0 \le m$ і $p_{\varphi(m, l, \dots, l)} = p_N = 2^{-(N-1)}$.

Тоді, аналогічно як в [5], для

$$E_k = \left(\frac{2^{ml^{k-1}}}{2^{ml^{k-1}} - 1} - \frac{l}{2^{ml^k} - 1}\right)(1 - 2^{-N}) + \frac{l}{2^N},$$

для $k = 0$

$$E_0 = \left(2 - \frac{m}{2^m - 1}\right)(1 - 2^{-N}) + \frac{m}{2^N}$$

і для $k = r$

$$E_r = \frac{2^{ml^{r-1}}}{2^{ml^{r-1}} - 1}(1 - 2^{-N}).$$

Тому, нехтуючи величиною 2^{-N}, з достатньо високою точністю можемо прийняти:

$$E = a_0 + ra_1 + \left(2 - \frac{m}{2^m - 1}\right)t_0 + \frac{2^{ml^{r-1}}}{2^{ml^{r-1}} - 1}t_1 + t_1\sum_{k=1}^{r-1}\left(\frac{2^{ml^{k-1}}}{2^{ml^{k-1}} - 1} - \frac{l}{2^{ml^k} - 1}\right).$$

Після перетворення

$$E = a_0 + ra_1 + 2t_0 - \frac{mt_0 - t_1}{2^m - 1} + \left(r - (l-1)\sum_{k=1}^{r-1}\frac{1}{2^{ml^k} - 1}\right)t_1.$$

Оскільки для будь-яких $m \ge 2, l \ge 2$

$$0 \le (l-1)\sum_{k=1}^{r-1}\frac{1}{2^{ml^k} - 1} \le (l-1)\sum_{k=1}^{r-1}\frac{1}{4^{lk} - 1} <$$

і $r \ge 1$, то $E \approx \bar{E}$, де:

$$\bar{E} = a_0 + r(a_1 + t_1) + 2t_0 - \frac{mt_0 - t_1}{2^m - 1}$$

або

$$\bar{E} = b_0 + d_0 m + \frac{b_1 + d_1 l + t_1}{\ln l}(\ln N - \ln m) + 2t_0 - \frac{mt_0 - t_1}{2^m - 1}.$$

Для наближеного визначення значень параметрів m і l, за яких E досягає мінімуму, маємо систему рівнянь:

$$\begin{cases} \ln l = 1 + \dfrac{1}{l}\dfrac{b_1 + t_1}{d_1}, \\ \dfrac{1}{(2^m - 1)^2}\left((2^m(m\ln 2 - 1) + 1)\dfrac{t_0}{d_0} - \dfrac{t_1}{d_0}2^m \ln 2\right) + 1 = \dfrac{1}{m\ln l}\left(\dfrac{b_1 + t_1}{d_0} + \dfrac{d_1}{d_0}l\right). \end{cases} \quad (2)$$

Зокрема, якщо $a_0 = const, a_1 = const$, то функція E досягає мінімуму, якщо $r = 1$. Це випливає з того, що:

$$f(r, m, l) = ra_1 + \left(r - (l-1)\sum_{k=1}^{r-1}\frac{1}{2^{ml^k} - 1}\right)t_1,$$

для довільних $m \ge 2, l \ge 2$ зростає з ростом кількості рівнів $r\ (r = 1, 2, \dots)$. Якщо $r = 1$, то отримуємо вираз:

$$E = a_0 + a_1 + 2t_0 + t_1 - \frac{mt_0 - t_1}{2^m - 1}$$

і для визначення значення параметра m , за якого E досягає мінімуму, маємо рівняння:

$$m \ln r + 2^{-m} = 1 + \frac{t_1}{t_0}\ln 2 .$$

Таблиця 3

Розв'язки системи (2) залежно від $\frac{b_1}{t_1}$ та $\frac{d_1}{t_1}$ у випадку $t_0 = t_1$, $\frac{d_0}{t_0} = 1000$ і $N = 10^6$

$\frac{b_1}{t_1}$	$\frac{d_1}{t_1}$		
	10	100	1000
10	$\tilde{l} = 3{,}07; m = 1$	$\tilde{l} = 2{,}83; m = 1$	$\tilde{l} = 2{,}73; m = 2{,}73$
100	$\tilde{l} = 8{,}69; m = 1$	$\tilde{l} = 3{,}60; m = 1$	$\tilde{l} = 2{,}82; m = 2{,}82$
1000	$\tilde{l} = 37{,}96; m = 1$	$\tilde{l} = 8{,}65; m = 1$	$\tilde{l} = 3{,}59; m = 3{,}59$

Якщо , то корінь отриманого рівняння є в межах ; якщо – в межах $2 < m < 2{,}2$; якщо – в межах $0 < m < 1$.

Отже, якщо враховувати, що $m \geq 2$ і m – ціле число, то в разі $a_0 = const$ і $a_1 = const$ матеметичне сподівання загального часу, необхідного для пошуку запису у файлі, досягає мінімуму у випадку, коли індекс має лише один рівень і файл розбитий по два записи у кожному.

Таблиця 4

Оптимальні значення $\frac{\bar{E}}{d_0}$ залежно від $\frac{b_1}{t_1}$ та $\frac{d_1}{t_1}$ у випадку $t_0 = t_1$, $\frac{d_0}{t_0} = 1000$, $\frac{b_0}{d_0} = 1000$ і $N = 10^6$

$\frac{b_1}{t_1}$	$\frac{d_1}{t_1}$		
	10	100	1000
10	2508,84	5906,50	38696,81
100	3202,61	6974,11	39825,80
1000	7247,02	13950,57	49624,75

Розподіл імовірностей звертань до записів за законом Зіпфа

Нехай ймовірності звертання до записів розподілені за законом Зіпфа [7], тобто:

$$p_i = \frac{1}{iH_i} \ (i = 1, 2, \ldots, N), H_N = \sum_{k=1}^{N} \frac{1}{k}.$$

Тоді:

$$p_{\varphi(i_0, i_1, \ldots, i_r)} = \frac{1}{\varphi(i_0, i_1, \ldots, i_r) H_N}$$

і, аналогічно як в [5], для $k = 1, 2, \ldots, r-1$

$$E_k = \frac{1}{H_N}\left(l S_{ml^k}\left(l^{r-k}\right) - S_{ml^{k-1}}\left(l^{r-k+1}\right) + H_N\right),$$

для $k = 0$

$$E_0 = \frac{1}{H_N}\left(m S_m\left(l^r\right) - N\left(H_N - 1\right)\right)$$

і для $k = r$

$$E_r = \frac{1}{H_N}\left((l+1)H_N - S_{ml^{r-1}}(l)\right),$$

де

$$S_{ml^i}\left(l^{r-i}\right) = \sum_{k=1}^{l^{r-i}} H_{kml^i}, \qquad i = 0, 1, \ldots, r-1.$$

Тому:

$$E = a_0 + r a_1 + \frac{1}{H_N}\left(\left(m S_m\left(l^r\right) - N\left(H_N - 1\right)\right)t_0 + \left((l + r)H_N + (l-1)\sum_{i=1}^{r-1} S_{ml^i}\left(l^{r-i}\right) - S_m\left(l^r\right)\right)t_1\right).$$

Представимо $S_{ml^i}\left(l^{r-i}\right)$ $(i = 0, 1, \ldots, r-1)$ у вигляді функції від l^{r-i}. Для цього використаємо апроксимацію часткової суми гармонічного ряду натуральним логарифмом. Отримаємо:

$$S_{ml^t}\left(l^{r-t}\right)=\sum_{k=1}^{l^{r-t}}\left(\ln kml^t+C+\gamma_{kml^t}\right)=\ln\left(l^{r-t}\right)!+l^{r-t}\ln ml^t+l^{r-t}\left(C+\bar{\gamma}_{kml^t}\right),$$

де $C=0{,}577\ldots$ – стала Ейлера, γ_{kml^t} $\left(k=1,2,\ldots,l^{r-t}\right)$ – деякі безмежно малі величини,

$$\bar{\gamma}_{ml^t}=\frac{1}{l^{r-t}}\sum_{k=1}^{l^{r-t}}\gamma_{kml^t}\,.$$

Враховуючи формулу Стірлінга для $\left(l^{r-t}\right)!$, отримаємо:

$$S_{ml^t}\left(l^{r-t}\right)=\frac{1}{2}\ln l^{r-t}+l^{r-t}\left(\ln l^{r-t}-1\right)+C_1+\frac{\theta}{12\,l^{r-t}}+l^{r-t}\ln ml^t+l^{r-t}\left(C+\bar{\gamma}_{ml^t}\right),$$

де $C_1=0{,}5\ln 2\pi,\ 0<\theta<1$. Оскільки $ml^r=N$, то:

$$S_{ml^t}\left(l^{r-t}\right)=l^{r-t}\left(\ln N+C+\bar{\gamma}_{ml^t}-1\right)+\frac{1}{2}\ln l^{r-t}+C_1+\frac{\theta}{12\,l^{r-t}}.$$

Нехтуючи доданком $\frac{\theta}{12\,l^{r-t}}$ і замінюючи $\ln N+C+\bar{\gamma}_{ml^t}$ значенням H_N, отримуємо:

$$S_{ml^t}\left(l^{r-t}\right)\approx l^{r-t}\left(H_N-1\right)+0{,}5\ln l^{r-t}+C_1.$$

Підставляючи в формулу для E замість функції $S_{ml^t}\left(l^{r-t}\right)$ її апроксимацію, дістаємо $E\approx\bar{E}$, де

$$\bar{E}=b_0+\frac{d_0N}{l^r}+r\left(b_1+d_1l\right)+\frac{1}{H_N}\left(\frac{Nt_0}{2l^r}\ln 2\pi l^r+\left(rH_N+l+\frac{1}{4}(l-1)(r-1)\ln 4\pi^2l^r-\frac{1}{2}\ln 2\pi l^r\right)t_1\right).$$

Для наближеного визначення значень параметрів r і l, за яких E досягає мінімуму, маємо систему рівнянь:

(3)

Зокрема, якщо $a_0=const,\ a_1=const$, то система рівнянь матиме вигляд:

$$\begin{cases}\dfrac{l}{r}\left(1+\dfrac{1}{4}(r-1)\ln 4\pi^2l^r\right)+\dfrac{1}{4}\left((r-1)(l-1)-2\right)=\dfrac{Nt_0}{2l^rt_1}\left(\ln 2\pi l^r-1\right),\\ l\ln l+\dfrac{1}{4}\left(l(r-1)\ln l-r(l-1)\right)\ln 4\pi^2l^r=rH_N\left(1+\dfrac{a_1}{t_1}\right).\end{cases}$$

Таблиця 5

Розв'язки системи (3) залежно від $\frac{b_1}{t_1}$ та $\frac{d_1}{t_1}$ у випадку $t_0=t_1,\ \frac{d_0}{t_0}=1000$ і $N=10^6$

$\frac{b_1}{t_1}$	$\frac{d_1}{t_1}$		
	10	100	1000
10	$l=3{,}62;r=13{,}28$	$l=2{,}82;r=14{,}53$	$l=2{,}73;r=12{,}76$
100	$l=8{,}55;r=7{,}57$	$l=3{,}59;r=11{,}59$	$l=2{,}82;r=12{,}34$
1000	$l=37{,}32;r=4{,}08$	$l=8{,}64;r=6{,}47$	$l=2{,}59;r=9{,}81$

Таблиця 6

Оптимальні значення $\frac{\bar{E}}{d_0}$ залежно від $\frac{b_1}{t_1}$ та $\frac{d_1}{t_1}$ у випадку $t_0=t_1,\ \frac{d_0}{t_0}=1000,\ \frac{b_0}{d_0}=1000$ і $N=10^6$

$\frac{b_1}{t_1}$	$\frac{d_1}{t_1}$		
	10	100	1000
10	1676,09	5552,06	38702,01
100	2517,09	6708,52	39831,11
1000	7030,71	13952,33	49630,95

На рис. 1 показана залежність $\frac{E}{d_0}$ від зміни закону розподілу ймовірностей звертання до записів для $\frac{b_1}{t_1}=100,\ \frac{d_0}{t_0}=1000$, $\frac{b_0}{d_0}=1000,$ $t_0=t_1$, $N=10^6$ і різних $\frac{d_1}{t_1}$.

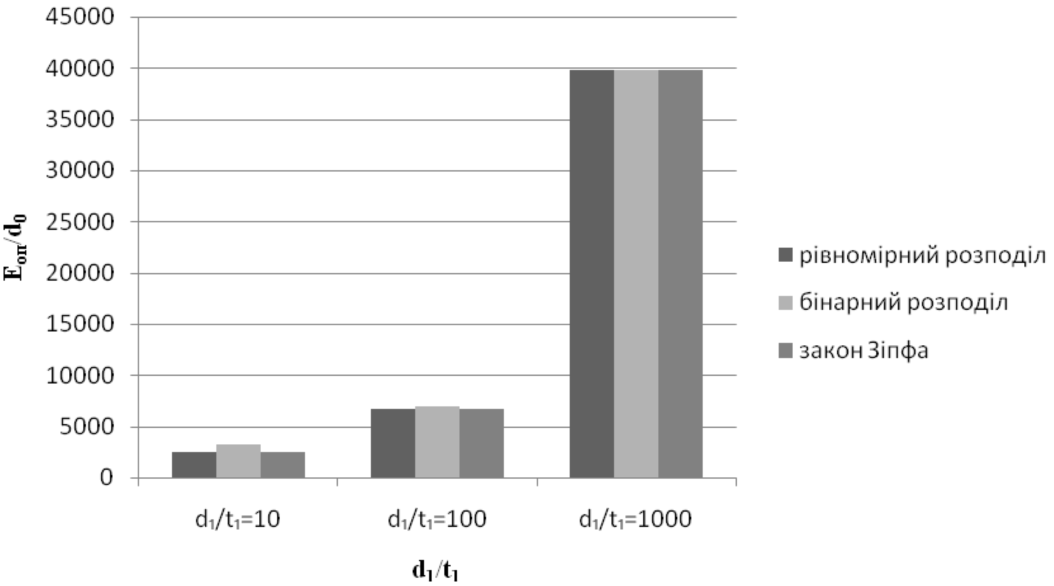

Рис. 1. Залежність $\frac{E}{d_0}$ від зміни закону розподілу ймовірностей звертань до записів

Висновки. Побудовано оптимальні схеми пошуку записів в індексо-послідовних файлах, які зберігаються у зовнішній пам'яті ЕОМ, для таких законів розподілу ймовірностей звертання до записів, як рівномірний, "бінарний", Зіпфа. 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. Як видно з порівняння, математичне сподівання загального часу, необхідного для пошуку запису в файлі, майже не залежить від розподілу ймовірностей звертань до записів.

Література

1. Мартин Дж. Организация баз данных в вычислительных системах / Мартин Дж. – М. : Мир, 1980. – 664 с.
2. Цегелик Г. Г. Оптимальные по времени поиска модели индекно-последовательных файлов при неравномерном распределении вероятностей обращения к записям / Г. Г. Цегелик // Программирование. – 1986. – № 2. – С. 81–86.
3. Цегелик Г. Г. Определение параметров оптимальной организации многоуровневых индексно-последовательных файлов / Г. Г. Цегелик // Кибернетика. – 1988. – № 2. – С. 74–78.
4. Цегелик Г. Г. Организация и поиск информации в базах даных / Цегелик Г. Г. – Львов : Вища школа, 1987. – 176 с.
5. Цегелик Г. Г. Системы распределенных баз даных / Цегелик Г. Г. – Львов : Свит, 1990. – 168 с.
6. Цегелик Г. Г. Моделювання та оптимізація доступу до інформації файлів баз даних для однопроцесорних і багатопроцесорних систем / Цегелик Г. Г. – Львів : Вид-во ЛНУ імені Івана Франка, 2010. – 192 с.
7. Кнут Д. Искусство програмирования для ЭВМ : Т. 3. Сортировка и поиск / Кнут Д. – М. : «Вильямс», 2007. – 824 с.

Надійшла 10.11.2010 р.