DOI 10.31891/2307-5732-2019-273-3-204-207
УДК 519.95

Л.І. ФУНДАК, Г.Г. ЦЕГЕЛИК
Львівський національний університет імені Івана Франка

ОПТИМАЛЬНИЙ БЛОКОВИЙ ПОШУК У ВИПАДКУ РІВНОМІРНОГО РОЗПОДІЛУ ЙМОВІРНОСТЕЙ ЗВЕРТАННЯ ДО ЗАПИСІВ

У роботі розглянутий оптимальний блоковий пошук у випадку рівномірного розподілу ймовірностей звертання до записів, якщо в локалізованому блоці використовується метод двійкового пошуку. Проведено дослідження оптимальної кількості порівнянь, необхідних для пошуку запису у файлі. Показано ефективність методу двійкового пошуку у порівнянні з використанням методу послідовного перегляду для пошуку запису у локалізованому блоці.

Ключові слова: рівномірний розподіл ймовірностей звертання до записів, блоковий пошук, метод двійкового пошуку, метод послідовного перегляду.

L.I. FUNDAK, H.H. TSEHELYK
Ivan Franko Lviv National University

OPTIMAL BLOCK SEARCH IN THE CASE OF A UNIFORM DISTRIBUTION OF PROBABILITIES OF ACCESS TO RECORDS

The purpose of the article is to investigate the effectiveness of the optimal block search in the case of a uniform distribution of the probabilities of access to records if the localized block uses the binary search method. Among the methods for finding information in large databases, the block search method is most effective. The essence of this method is as follows. If file entries ordered in ascending or decreasing values of a key, then it is not necessary to view all records preceding the searched for the record. Entries can be split into blocks and first locate the block containing the desired entry by viewing the latest block records. After the record block is localized, the search for the desired record in the block is continued using one of the methods below. Block search method investigated for different laws of distribution of the probabilities of access to records when used in a localized block to search for the method of sequential viewing. However, in the case of a uniform distribution of probabilities, the block search method can be made much more efficient by using the binary search method in the localized block. This case investigated in the work. The graphs show the dependence of the average number of comparisons between the number of records in the file and the number of blocks on which the file is split. A comparison of the effectiveness of two block search options (with using in block sequential viewing and binary search) is conducted. The average number of comparisons for the different number of records in the file for both methods was calculated and compared. It is shown, that using a binary search method to search for a record in a localized block, you can significantly reduce the average number of comparisons required to search for a record in a file.

Keywords: uniform distribution of probabilities of access to records, block search, binary search method, sequential search.

Вступ

Якщо записи файлу впорядковані за зростанням чи спаданням значень ключа, то для пошуку запису не обов'язково переглядати всі записи, що передують шуканому. Записи можна розбити на блоки і спочатку локалізувати блок, який містить шуканий запис, переглядаючи останні записи блоків. Після того, як блок записів локалізований, пошук потрібного запису у блоці продовжують за допомогою одного з розглянутих нижче методів.

В [1] побудовані оптимальні моделі блокового пошуку для різних законів розподілу ймовірностей звертання до записів у разі використання методу послідовного перегляду для пошуку запису у локалізованому блоці. Однак, у випадку рівномірного розподілу ймовірностей звертання до записів можна значно зменшити середню кількість порівнянь, необхідних для пошуку запису у файлі, якщо в локалізованому блоці використати метод двійкового пошуку.

Основна частина

Нехай усі записи впорядкованого файлу розбиті на n блоків по m записів у кожному, тобто кількість записів у файлі рівна $N = nm$. Вважатимемо, що розподіл ймовірностей звертання до записів є рівномірний. Тоді при використанні в локалізованому блоці методу послідовного перегляду середня кількість порівнянь, необхідних для пошуку запису у файлі, виражається формулою [1]

$$E' = \frac{1}{2}(n+m)+1.$$

Функція E' є опуклою та існує єдина точка, у якій вона досягає мінімуму. У роботі [1] показано, що оптимальне значення функції досягається при $n = m = \sqrt{N}$.

Отже, оптимальне число порівнянь E', необхідних для пошуку запису у файлі, обчислюватимемо за такою формулою

$$E'_{on} = \sqrt{N}+1. \quad (1)$$

Нехай тепер кількість записів у впорядкованому файлі $N = (2^l - 1)n$, де n – кількість блоків, на які розбитий файл, а $(2^l - 1)$ – кількість записів у кожному блоці. Вважатимемо, що розподіл ймовірностей звертання до записів є рівномірний. Тоді при використанні в локалізованому блоці методу двійкового

пошуку середня кількість порівнянь, необхідних для пошуку запису у файлі, виражається формулою [2]

$$E'' = \frac{1}{2}(n+1) + \frac{1}{2^l - 1}\sum_{i=1}^{l} i\, 2^{i-1},$$

Оскільки [2]

$$\sum_{i=1}^{l} i\, 2^{i-1} = (l-1)2^l + 1,$$

то

$$E'' = \frac{1}{2}(n+1) + l - 1 + \frac{l}{2^l - 1}.$$

Знайдемо значення параметрів n і l, за яких середня кількість порівнянь E'' досягає мінімуму.

Надамо E'' у вигляді функції від однієї змінної. Для цього зробимо заміну змінних $n = \frac{N}{2^l - 1}$, одержимо

$$E'' = \frac{1}{2^l - 1}\left(\frac{N}{2} + l\right) + l - \frac{1}{2}. \quad (2)$$

Для знаходження точки екстремуму функції (2) дістаємо рівняння [3]

$$2^l - l\ \ln 2 = 1 + \frac{1}{2} N \ln 2. \quad (3)$$

У [3] показано, що одержане рівняння має єдиний додатний корінь $l = l_o$, і цей корінь є точкою мінімуму опуклої функції $E'' = E''(l)$.

У табл. 1 наведені корені l_o рівняння (3) і обчислені за формулою (2) значення E''_{on} для різних N.

Таблиця 1

Оптимальне значення E''_{on} для різних N

l	$2^l - 1$	n	N	l_o	E''_{on}
1	1	10	10	2,65706	3,59976
2	3	10	30	3,81137	4,75407
2	3	50	150	5,83364	6,77633
2	3	100	300	6,77702	7,71971
3	7	10	70	4,83865	5,78135
3	7	50	350	6,99034	7,93304
3	7	100	700	7,96069	8,90339
4	15	10	150	5,83364	6,77633
4	15	50	750	8,05808	9,00077
4	15	100	1500	9,04201	9,9847
5	31	10	310	6,82231	7,765
5	31	50	1550	9,08876	10,0315
5	31	100	3100	10,08	11,0227
6	63	10	630	7,81222	8,75491
6	63	50	3150	10,1029	11,0456
6	63	100	6300	11,0981	12,0408
7	127	10	1270	8,80493	9,74764
7	127	50	6350	11,1095	12,0522
7	127	100	12700	12,1068	13,0495

На рис. 1, 2 показано поведінку l_o і E''_{on} для різних значень N.

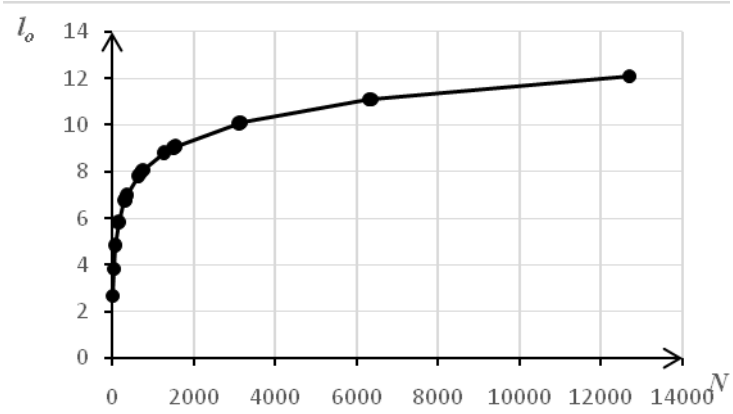

Рис. 1. Поведінка l_o для різних значень N

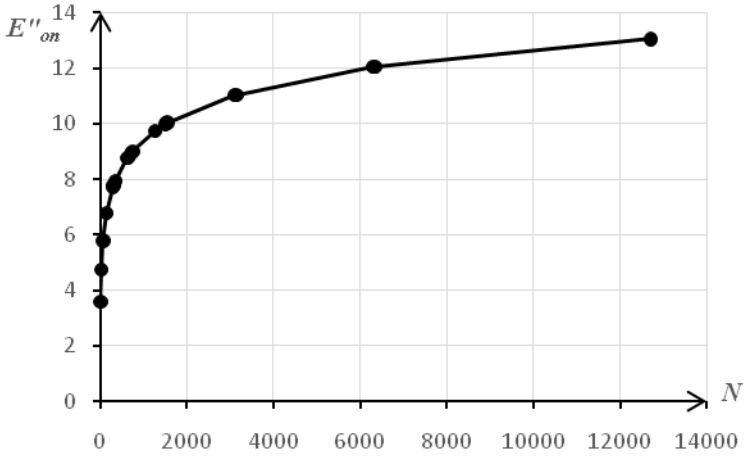

Рис. 2. Поведінка E''_{on} для різних значень N

На рис. 3 зображено графік поведінки функції E'' в околі точки мінімуму для $N = 300$, при якому оптимальне значення $l_0 \approx 6,777$, а $E''_{on} = 7,719$.

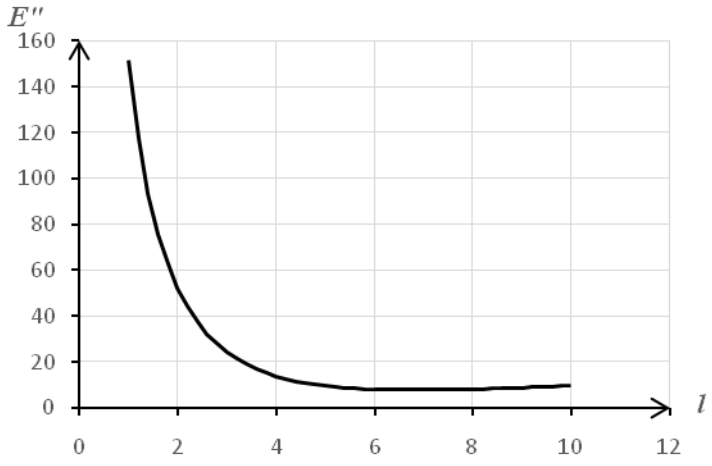

Рис. 3. Графік поведінки E'' в околі точки мінімуму для $N = 300$

Проведемо порівняльний аналіз оптимального значення числа порівнянь, необхідних для пошуку запису у файлі, у випадку використання в локалізованому блоці методу послідовного перегляду (E'_{on}) і методу двійкового пошуку (E''_{on}) для різних значень N . У табл. 2 наведено результати обчислень для обох методів.

Таблиця 2

Оптимального значення E'_{on} і E''_{on} для різних N

N	E'_{on}	E''_{on}
150	13,24745	6,77633
300	18,32051	7,71971
700	27,45751	8,90339
1500	39,72983	9,9847
3100	56,67764	11,0227
6300	80,37254	12,0408
12700	113,6943	13,0495

На рис. 4 зображено графік поведінки функцій E'_{on} і E''_{on} для різних значень N.

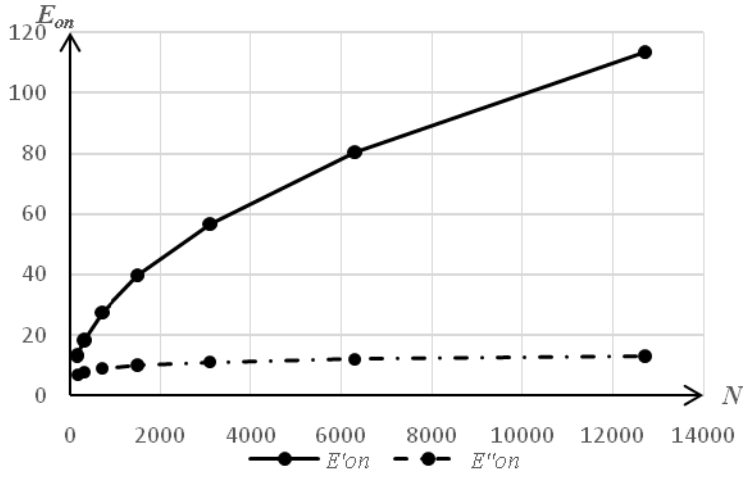

Рис. 4. Графік поведінки функцій E'_{on} і E''_{on}

Висновки

Проведено дослідження оптимальної кількості порівнянь, необхідних для пошуку запису в методі блокового пошуку при використанні в локалізованому блоці методу двійкового пошуку. Здійснено порівняльний аналіз блокового пошуку під час використання в локалізованому блоці методу послідовного перегляду і методу двійкового пошуку. Одержані результати показують, що у випадку рівномірного розподілу ймовірностей звертання до записів середня кількість порівнянь, необхідних для пошуку запису у файлі, буде значно меншою, якщо в локалізованому блоці використати метод двійкового пошуку.

Література

1. Цегелик Г.Г. Моделювання та оптимізація доступу до інформації файлів баз даних для однопроцесорних і багатопроцесорних систем : монографія / Г. Г. Цегелик. – Львів, 2010. – 192 с.

2. Цегелик Г. Г. Методы автоматической обработки информации / Г. Г. Цегелик. – Львов, 1981. – 132 с.

3. Фундак Л. І. Оптимальний блоковий пошук у випадку рівномірного розподілу ймовірностей звертання до записів / Л.І. Фундак, Г.Г. Цегелик // Матеріали XXIV Всеукр. наук. конф. “Сучасні проблеми прикладної математики та інформатики”. – Львів, 2018. – С. 166–168.

References

1. Tsehelyk H.H. Modeliuvannia ta optymizatsiia dostupu do informatsii failiv baz danykh dlia odnoprotsesornykh i bahatoprotsesornykh system : monohrafiia / H. H. Tsehelyk. – Lviv, 2010. – 192 s.

2. Cegelik G. G. Metody avtomaticheskoj obrabotki informacii / G. G. Cegelik. – Lvov, 1981. – 132 s.

3. Fundak L. I. Optymalnyi blokovyi poshuk u vypadku rivnomirnoho rozpodilu ymovirnostei zvertannia do zapysiv / L.I. Fundak, H.H. Tsehelyk // Materialy XXIV Vseukr. nauk. konf. “Suchasni problemy prykladnoi matematyky ta informatyky”. – Lviv, 2018. – S. 166–168.

Рецензія/Peer review : 15.5.2019 р. Надрукована/Printed : 2.6.2019 р.
Рецензент: д.т.н., проф. Притула М.М.