УДК 519.68

М.І.Філяк, Г.Г.Цегелик, Х.С.Дороцька
Львівський національний університет ім. І.Франка,
кафедра обчислювальної математики

ПОРІВНЯЛЬНИЙ АНАЛІЗ ЕФЕКТИВНОСТІ МЕТОДУ ПОСЛІДОВНОГО ПЕРЕГЛЯДУ ДЛЯ РІЗНИХ ЗАКОНІВ РОЗПОДІЛУ ЙМОВІРНОСТЕЙ ЗВЕРТАННЯ ДО ЗАПИСІВ

The analysis of effectiveness of the sequential searching method for various laws of probabilities distribution of requesting to records has been considered.

Основою сучасних інформаційних технологій є бази даних (БД) і системи керування базами даних (СКБД). Еволюція СКБД відбувається на тлі безпрецендентного зростання кількості різноманітних застосувань ЕОМ, а технологія БД в свою чергу забезпечує необхідний фундамент такого зростання. У випадку розв'язування різноманітних нечислових задач з використанням концепції БД основний акцент переноситься з процедур обробки інформації на процедури організації збереження та пошуку інформації в БД. Тому продуктивність інформаційних систем, ядром яких є величезні БД, в значній мірі визначається ефективністю методів пошуку інформації в файлах БД. На сучасному етапі розвитку інформаційних систем для пошуку інформації в файлах БД, як правило, використовується метод послідовного перегляду, однорівневий і багаторівневий блочний пошук та двійковий пошук. Дослідженню ефективності цих методів присвячена низка робіт, зокрема [1-5], однак повного і глибокого їх аналізу для різних законів розподілу ймовірностей звертання до записів немає. Потреба в проведенні такого аналізу випливає із того, що в більшості систем обробки інформації типовими є якраз випадки нерівномірного розподілу ймовірностей звертання до записів.

В роботі проводиться порівняльний аналіз ефективності методу послідовного перегляду для різних законів розподілу ймовірностей зверстання до записів. За критерій ефективності приймається математичне сподівання числа порівнянь, необхідних для пошуку запису в файлі.

Розглянемо послідовний файл, який містить N записів. Припустимо, що в файлі треба знайти запис із значенням ключа K. Суть методу послідовного перегляду полягає в тому, що значення K послідовно порівнюється із значеннями ключа записів, починаючи з першого, доти, доки два значення, що порівнюються, не співпадуть. Якщо p_i – ймовірність звертання до i-го запису файла, то математичне сподівання числа порівнянь, необхідних для пошуку запису в файлі, виразиться формулою

$$E = \sum_{i=1}^{N} i p_i .$$

Знайдемо явний вираз для E у випадку різних законів розподілу ймовірностей звертання до записів і дослідимо залежність математичного сподівання від числа записів файла та від зміни закону розподілу ймовірностей.

1. Нехай розподіл ймовірностей звертання до записів є рівномірним. Тоді

$$E=\frac{1}{2}(N+1).$$

2. Якщо ймовірність звертання до записів задовольняють "бінарний" розподіл, тобто

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

то

$$E=\sum_{i=1}^{N-1}\frac{i}{2^i}+\frac{N}{2^{N-1}}=\sum_{i=1}^{N}\frac{i}{2^i}+\frac{N}{2^N}.$$

Позначимо

$$E_m=\sum_{i=1}^{m-1}\frac{i}{2^i}+\frac{m}{2^{m-1}}.$$

Покажемо, що справедлива така рекурентна формула

$$E_m=\frac{1}{2}E_{m-1}+1\quad (m=3,4,...).$$

Справді,

$$E_{m+1}=\sum_{i=1}^{m}\frac{i}{2^i}+\frac{m+1}{2^m}=\frac{1}{2}\sum_{i=1}^{m}\frac{i}{2^{i-1}}+\frac{m+1}{2^m}=$$

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

$$=\frac{1}{2}\left(\sum_{i=1}^{m}\frac{1}{2^{i-1}}+\sum_{i=1}^{m-1}\frac{i}{2^i}\right)+\frac{m+1}{2^m}=\frac{1}{2}\left(2-\frac{1}{2^{m-1}}+E_m-\frac{m}{2^{m-1}}\right)+\frac{m+1}{2^m}=$$

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

Використовуючи рекурентну формулу, одержуємо

$$E_3=1+\frac{1}{2}E_2,$$

$$E_3 = 1+\frac{1}{2}E_3 = 1+\frac{1}{2}+\frac{1}{2^2}E_2,$$

. . . . . .

$$E_N = 1+\frac{1}{2}E_{N-1} = 1+\frac{1}{2}+\ldots+\frac{1}{2^{N-2}}E_2.$$

Оскільки $E_2 = 1+\frac{1}{2}$, то

$$E_N = 1+\frac{1}{2}+\frac{1}{4}+\ldots+\frac{1}{2^{N-1}} = 2-\frac{1}{2^{N-1}}.$$

Отже, для "бінарного" розподілу

$$E = 2-\frac{2}{2^N}.$$

3. Припустимо, що ймовірності звертання до записів задовольняють 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

$$E = \sum_{i=1}^{N}\frac{i}{iH_N} = \frac{N}{H_N}.$$

Оскільки $H_N = \ln N + C + \gamma_N$, де $C = 0{,}577\ldots$ — стала Ейлера, γ_N — деяка нескінченно мала величина, то

$$E = \frac{N}{\ln N + C + \gamma_N}.$$

Нехтуючи величиною γ_N, з достатньо високою точністю можемо прийняти

$$E = \frac{N}{\ln N + C}.$$

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

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

де $0<c<1$,

$$H_N^{(c)} = \sum_{k=1}^{N} \frac{1}{k^c}.$$

Тоді

$$E = \sum_{i=1}^{N} \frac{i}{i^c H_N^{(c)}} = \frac{1}{H_N^{(c)}} \sum_{i=1}^{N} \frac{1}{i^{c-1}} = \frac{H_N^{(c-1)}}{H_N^{(c)}}.$$

Використовуючи апроксимації [6]

$$H_n^{(c)} = \frac{1}{1-c} n^{1-c} - C^{(c)} + \gamma_n^{(c)},$$

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

де $C^{(c)}$ – деякі сталі, $\gamma_n^{(c)}$ – деякі нескінченно малі величини, $\alpha^{(c)}(n)$ – повільно зростаюча функція (для значень цієї функції складені таблиці), одержуємо

$$E = \frac{\frac{1}{2-c} N^{2-c} + \alpha^{(c)}(N)}{\frac{1}{1-c} N^{1-c} - C^{(c)} + \gamma_N^{(c)}}.$$

Нехтуючи нескінченно малою величиною $\gamma_N^{(c)}$, з достатньо високою точністю можна прийняти

$$E = \frac{\frac{1}{2-c} N^{2-c} + \alpha^{(c)}(N)}{\frac{1}{1-c} N^{1-c} - C^{(c)}}.$$

Залежність математичного сподівання числа порівнянь, необхідних для пошуку запису в файлі, від числа записів файла для різних законів розподілу ймовірностей звертання до записів приведена на рис.

Зауважимо, що на рис. графіку $c=0$ відповідає рівномірний розподіл ймовірностей звертання до записів, а $c=1$ — закон Зіпфа.

На діаграмі приведена залежність математичного сподівання числа порівнянь, необхідних для пошуку запису в файлі, від зміни закону розподілу ймовірностей звертання до записів ($N=10^6$).

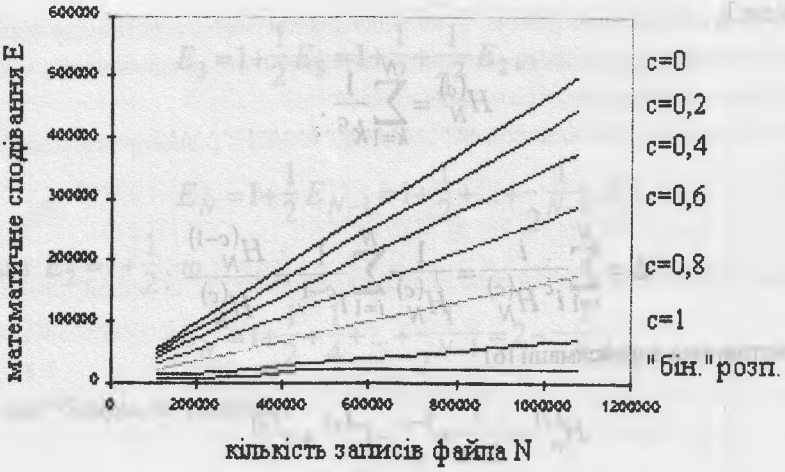

Рис. Залежність математичного сподівання числа порівнянь, необхідних для пошуку запису в файлі, від числа записів файла

Діаграма. Залежність математичного сподівання числа порівнянь, необхідних для пошуку запису в файлі, від зміни закону розподілу ймовірностей звертання до записів ($N = 10^6$)

1. Кнут Д. Искусство программирования для ЭВМ. Т.3. Сортировка и поиск. М.: Мир. 1978.-844 с. 2. Мартин Дж. Организация баз данных в вычислительных системах. М.: Мир. 1980.-644 с. 3. Філяк М.І., Цегелик Г.Г. Ефективність методів послідовного перегляду і блочного пошуку для різних законів розподілу ймовірностей звертання до записів// Вісн.Львів.у-ту. Сер.мех.-мат. 1998. Вип.50. С.200-203. 4. Цегелик Г.Г. Методы автоматической обработки информации. Львов: Вища шк. 1981.-132с. 5. Shneiderman B. Jump searching: a fast seguential search technigue// Comm.ACM. 1978. Vol.21. №10. P.831-

834. 6. Цегелик Г.Г., Кудеравець Х.С. Апроксимація часткових сум узагальнених гармонічних рядів неперервними функціями // Вісн.Львів.у-ту. Сер.мех.-мат. 1995. Вип.42. С. 17-19.

УДК 681.3

В.С.Якушев, Д.А.Чепурний

НУ "Львівська політехніка", кафедра інформаційних систем та мереж

ПЕРЕТВОРЕННЯ Й ОПРАЦЮВАННЯ ІНФОРМАЦІЇ З ВИКОРИСТАННЯМ АЛГОРИТМУ КАНТОРА І ТЧП

The method for number-theoretic transformations (NTT) with using residual class number system is presented. It is based on number representation of Cantor algorithm. The result is a computing complication and hardware cost reduction of NTT.

Застосування алгоритму Кантора для перетворення і опрацювання інформаційних потоків вважається ефективним у випадку подальшого цифрового опрацювання сигналів (електричних) за допомогою апарату теоретико-числових перетворень, зокрема Китайської теореми про залишки (КТЗ) [1]. Алгоритм Кантора [2] дозволяє відображати вхідну аналогову величину в цифровому вигляді, який найприйнятніший для застосування КТЗ, тобто у вигляді залишків за кількома модулями. КТЗ підвищує швидкодію опрацювання інформації за рахунок здійснення обчислень паралельно і незалежно за кожним з модулів. Окрім того, сумісне застосування алгоритму Кантора і КТЗ підвищує швидкодію переходу від числового відображення в нормальному вигляді до системи залишків і навпаки, а також знижує складність цього переходу, що є одним з основних стримуючих факторів для застосування при обчисленні дійсних чисел у вигляді системи залишків [3]. Ця стаття є продовженням роботи [2].

Метою досліджень є зменшення обчислювальної складності та апаратурної вартості обчислення теоретико-числових перетворень (ТЧП) за рахунок використання відображення дійсних чисел швидкозбіжними рядами, а також отримання співвідношень, які дозволяють оцифровувати дані в необхідному для ТЧП вигляді на основі сучасних способів перетворення інформації, в тому числі аналого-цифрового перетворення.

Основними методами, що використовуються при цифровому опрацюванні сигналів, є згортка та гармонічний аналіз. Відомо, що згортку можна обчислювати за допомогою спектральних методів, але порівняно недавно стало відомо, що й гармонічний аналіз можна проводити, обчислюючи систему згорток.

Поява нових порівняно з швидким перетворенням Фур'є [1] алгоритмів, таких як алгоритми простих множників Томаса-Гуда, алгоритм Рейдера (зведення до циклічної