Дискретная математика

том 7 выпуск 3 * 1995

УДК 519.2

Некоторые оценки для распределения высоты дерева цифрового поиска

© 1995 г. В. А. Ватутин, В. Г. Михайлов

Пусть $\varkappa(T)$ — высота q-арного дерева цифрового поиска T, построенного по ключам $K_1, \dots, K_n$, каждый из которых является вектором с компонентами, принадлежащими алфавиту $\mathfrak{A} = \{0, 1, \dots, q-1\}$. В предположении, что компоненты векторов независимы и равномерно распределены на $\mathfrak{A}$, найдены оценки сверху и снизу для вероятностей $\mathsf{P}\{\varkappa(T) \leqslant m\}$, $m = 1, \dots, n$, с явно выписанными константами. Для типичных значений m полученные соотношения являются асимптотически более точными по сравнению с утверждением, доказанным в [2].

Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований, проект 93–011–1443.

Необходимость обработки больших массивов данных на ЭВМ привела к созданию различных способов хранения и поиска информации. Одним из таких способов является метод цифрового поиска, состоящий в следующем. Каждой единице хранения сопоставляется ключ $K = (k^1, \dots, k^l)$, представляющий собой вектор из цифр (букв) некоторого алфавита $\mathfrak{A}$. Таким образом, вся хранящаяся информация описывается определенным множеством ключей $\{K_1, \dots, K_n\}$. Этому множеству сопоставляется граф, являющийся корневым деревом, ребра которого помечены буквами алфавита $\mathfrak{A}$, а висячие вершины (листья) соответствуют некоторым наборам ключей множества $\{K_1, \dots, K_n\}$. При этом должен выполняться ряд условий. Во-первых, наборы ключей на листьях не пересекаются и в сумме дают множество ключей $\{K_1, \dots, K_n\}$. Во-вторых, начальные знаки (буквы) ключей, принадлежащих одному листу, повторяют последовательность меток на ребрах, образующих путь от корня к данному листу.

В качестве примера рассмотрим алфавит $\mathfrak{A} = \{0, 1\}$ и набор ключей $\{001, 0001, 111, 10000, 10001, 10110\}$. Указанными свойствами обладают деревья, указанные на рис. 1.

При соответствующем способе размещения ключей и информации в памяти ЭВМ дерево $T(K_1, \dots, K_n)$ может быть использовано для извлечения информации из массива. Для проверки того, имеется ли искомый ключ K среди $K_1, \dots, K_n$, производится поразрядное сравнение знаков ключа K с метками на ребрах дерева и (при совпадении очередного разряда с соответствующей меткой) продвижение по ребрам от корня к листьям. В результате либо обнаруживается список (лист), содержащий ключ K, либо устанавливается отсутствие этого ключа среди $K_1, \dots, K_n$.

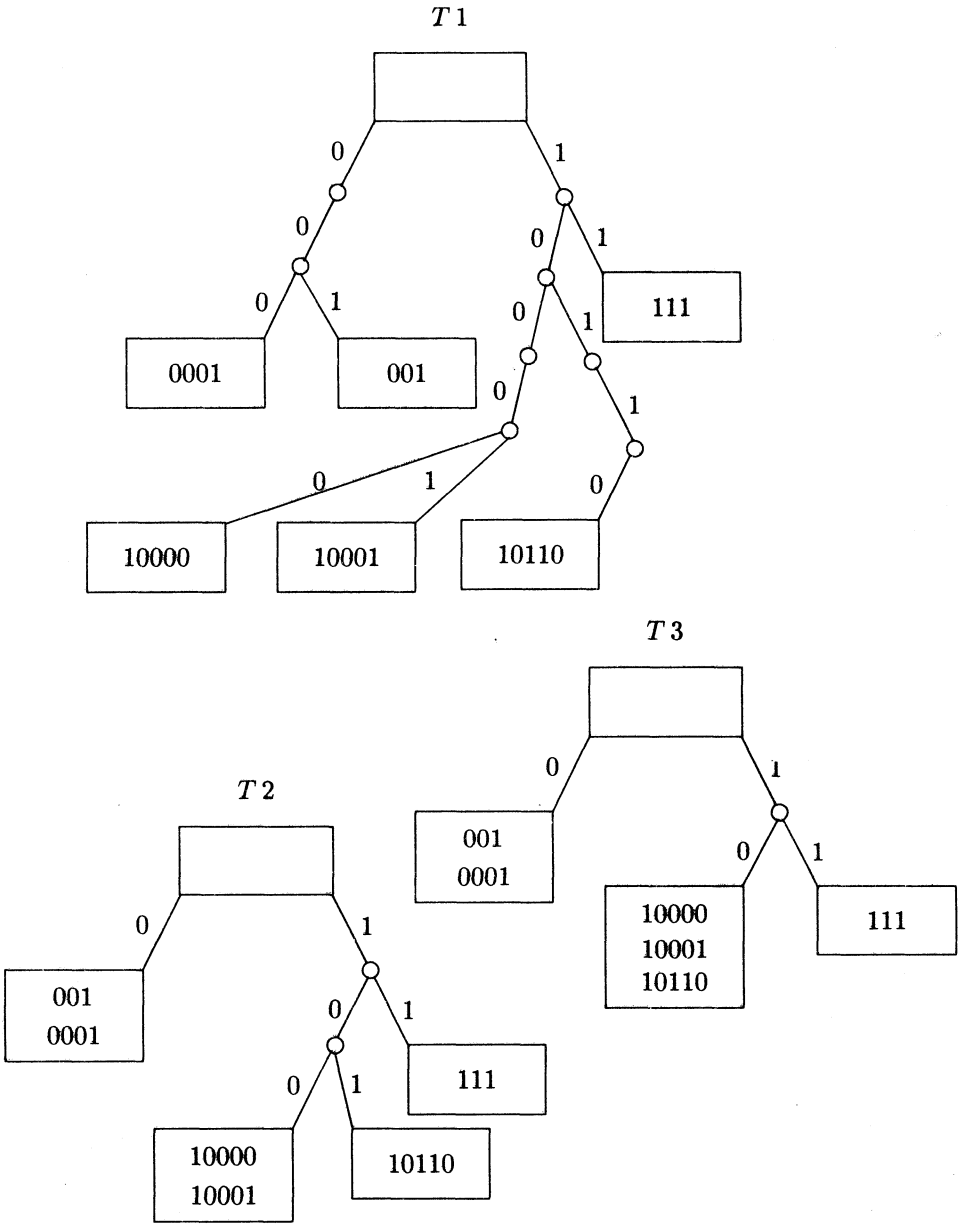
T 1
0
1
0
0
1
0
1
0
1
0
1
0
1
0
0
1
0
1
0
0001
001
111
10000
10001
10110
T 3
0
1
0
1
001
0001
10000
10001
10110
111
T 2
0
1
0
1
0
1
001
0001
111
10000
10001
10110

Рис. 1.

В зависимости от вида обрабатываемой информации описанная процедура организации хранения и поиска данных может быть лучше или хуже (с различных точек зрения) других распространенных способов обработки информации, например, бинарного поиска в упорядоченном массиве. Подробное обсуждение такого рода проблем имеется в [1].

Среди характеристик, описывающих условия хранения и поиска информации с использованием дерева $T(K_1, \dots, K_n)$ наиболее важными являются высота дерева $\varkappa(T(K_1, \dots, K_n))$, оценивающая сверху время, затрачиваемое на успешный или безуспешный поиск ключей; объем памяти, которую необходимо зарезервировать для построения дерева $T(K_1, \dots, K_n)$ и соответствующей организации хранения ключей.

Вторая характеристика тесно связана с первой. Поскольку априори конкретный вид дерева поиска неизвестен, необходимо предусмотреть сразу все возможности, что достигается резервированием памяти для размещения полного $|\mathfrak{A}|$-арного дерева высоты $\varkappa(T(K_1, \dots, K_n))$. В связи с этим требования к объему памяти должны выражаться в терминах величин $|\mathfrak{A}|^{\varkappa}$ и $\varkappa$. Свойства этих величин мы и изучаем в данной заметке.

Далее нам необходимо договориться о терминологии и обозначениях. В качестве алфавита мы будем рассматривать множество $\mathfrak{A} = \{0, \dots, q-1\}$, $q \geqslant 2$. Дерево цифрового поиска, построенное на основе цифр алфавита $\mathfrak{A}$, мы будем называть, согласно терминологии [1, гл. 6.3], q-арным бором. Хотя, на наш взгляд, это название не совсем удачно, частичным оправданием его служит тот факт, что корень бора можно расщепить и интерпретировать его как совокупность q корней, занумерованных числами $0, \dots, q-1$, и, следовательно, соответствующие этим корням деревья (некоторые из них могут быть пустыми) образуют лес, деревья в котором имеют густую крону, если, например, q велико, длина всех ключевых последовательностей одна и та же, скажем l, а общее количество ключей близко к q^l.

Обратим внимание на некоторые различия деревьев $T1$, $T2$ и $T3$, изображенных на рис. 1. В первом случае ветвление остановилось лишь тогда, когда на каждый лист стало приходиться не более одного ключа. Во втором случае путь заканчивался листом, как только на нем оставалось не более двух ключей, и, наконец, в третьем ветвление прекращалось, если оставалось не более трех ключей. В первом случае мы называем полученное дерево 1-бором, во втором 2-бором и в третьем 3-бором. Аналогично вводится понятие b-бора при $b \geqslant 4$.

Мы будем считать, что знаки, образующие ключи $K_1, \dots, K_n$, выбраны случайно, независимо и равновероятно из алфавита $\mathfrak{A}$, длины ключей выбраны так, чтобы ключи различались (вероятность этого события, очевидно, равна единице, поскольку длина ключевых последовательностей не ограничивается) и все лишние знаки в правых разрядах ключей отброшены. Последняя процедура по сути дела эквивалентна построению 1-бора с последующим вычеркиванием тех правых разрядов ключевых последовательностей, которые не участвовали в построении. Поскольку множество ключей такого рода однозначно определяет дерево поиска (b-бор), если принять соглашение о том, что ребра, выходящие из вершин, нумеруются в порядке возрастания значений меток слева направо, то однозначно определена и случайная величина $\varkappa_q(n, b)$ — высота этого дерева (b-бора).

Закон распределения высоты случайного q-арного b-бора с n листьями опи-

сывается вероятностями

$$P_{q,n,b}(m) = \mathbf{P}\{\varkappa_q(n,b) = m\}$$

или, что одно и то же, вероятностями

$$F_{q,n,b}(m) = \mathbf{P}\{\varkappa_q(n,b) \leqslant m\}.$$

Очевидно,

$$F_{q,n,b}(m) = \sum_{k=1}^{m} P_{q,n,b}(k). \tag{1}$$

При небольших значениях параметров n, b и m вычислить значения величин $F_{q,n,b}(m)$ нетрудно по формуле

$$F_{q,n,b}(m) = n!\,(q^m)!\,q^{-mn} \sum \left(\prod_{r=0}^{b} ((r!)^{m_r} m_r!) \right)^{-1}, \tag{2}$$

где суммирование проводится по множеству наборов неотрицательных чисел $(m_0, \ldots, m_b)$, удовлетворяющих условиям

$$\sum_{r=0}^{b} m_r = q^m, \qquad \sum_{r=1}^{b} r m_r = n. \tag{3}$$

Доказательство формулы (2) будет дано позднее.

С ростом n сложность вычислений по формуле (2) и другим подобным ей формулам быстро растет, что вызывает необходимость нахождения аппроксимирующих соотношений. Первый шаг в получении таких формул был сделан в работе [2] (см. также статьи [3, 4]).

Теорема [2]. *Пусть*

$$m \geqslant \left(1 + \frac{1}{b}\right) \log_2 n - \log_2 \log_2 n. \tag{4}$$

Тогда

$$F_{2,n,b}(m) = \exp\left\{ -\frac{1}{(b+1)!} 2^{(b+1)\log_2 n - bm} \right\} \left(1 + O\left(\frac{(\log_2 n)^{b+1}}{n^{1/b}} \right) \right). \tag{5}$$

Эта оценка дает представление о характере асимптотического поведения вероятностей $F_{2,n,b}(m)$, однако она совершенно непригодна для численных оценок при конкретных значениях параметров. Поэтому в настоящей статье мы делаем следующий шаг — строим явные оценки для вероятностей $F_{q,n,b}(m)$. Они пригодны для реального оценивания уже при умеренных значениях параметра n. Более того, используя эти оценки, можно получить асимптотическое представление для вероятностей $F_{q,n,b}(m)$, которое в типичных случаях является более точным, нежели равенство (5).

В последующих формулах мы полагаем $(a)_+ = \max(0,a)$, $0^{-1} = \infty$. Пусть для $m = 1, 2, \ldots$

$$\alpha_b(n,m) = (b+1)^2 \binom{n}{b+1} n^{-1} q^{-bm} \left(2\binom{n}{b+1} q^{-bm} + b + 4\right),$$

$$C_b(n,m) = 2(b+1)^2 \binom{n}{b+1}^2 n^{-1} q^{-2bm},$$

$$D_b(n,m) = (b+1)(b+2)\binom{n}{b+1} q^{-(b+1)m} \exp\left\{\frac{1}{2}(b+1)nq^{-m}\right\},$$

$$A_1(n,m) = 1 - \binom{n}{3} q^{-2m} - 2(1 + 2\alpha_1(n,m))C_1(n,m)\left(1 + nq^{-m}\right),$$

а при $b \geqslant 2$

$$A_b(n,m) = 1 - \binom{n}{b+2} q^{-(b+1)m} - 2(1 + 3\alpha_b(n,m))C_1(n,m),$$

$$B_b(n,m) = (1 - (1 + 2\alpha_b(n,m))D_b(n,m) - C_b(n,m))_+^{-1}.$$

Теорема 1. *Пусть $n > b + 1 \geqslant 2$, выполнено неравенство*

$$\alpha_b(n,m) \leqslant e^{-1}, \tag{6}$$

и если $b \geqslant 2$, то

$$(b+1)nq^{-m} \leqslant 1. \tag{7}$$

Тогда

$$A_b(n,m) < \mathbf{P}\{\varkappa_q(n,b) \leqslant m\} \exp\left\{\binom{n}{b+1} q^{-bm}\right\} < B_b(n,m). \tag{8}$$

Следствие 1. *Пусть $n > b + 1 \geqslant 2$, $\alpha_b(n, m-1) \leqslant e^{-1}$, и, если $b \geqslant 2$, то $(b+1)nq^{1-m} \leqslant 1$. Тогда справедливы оценки*

$$\mathbf{P}\{\varkappa_q(n,b) = m\} \leqslant e^{-\gamma}\left(1 - e^{-\gamma}\right) A_b(n, m-1) + e^{-\gamma}(B_b(n,m) - A_b(n,m-1)), \tag{9}$$

$$\mathbf{P}\{\varkappa_q(n,b) = m\} \geqslant e^{-\gamma}\left(1 - e^{-\gamma}\right) B_b(n,m) + e^{-\gamma}(A_b(n,m) - B_b(n,m-1)), \tag{10}$$

где

$$\gamma = \binom{n}{b+1} q^{-bm}.$$

Следствие 2. *Пусть параметры n, b и k удовлетворяют условиям*

$$n \geqslant 2e(b+1)^{2b}(b+4)^2, \qquad k \geqslant 4 - \frac{1}{b(b+1)} \log_q n. \tag{11}$$

Тогда

$$\left|F_{q,n,b}\left(\frac{1}{b}\log_q\binom{n}{b+1}+k\right)\exp\left\{q^{-b(k-l(n,b))}\right\}-1\right| \leqslant 11(b+1)^{2+1/b}n^{-1/b}q^{-(b+1)(k-l(n,b))}, \quad (12)$$

где

$$l(n,b)=\left\{\frac{1}{b}\log_q\binom{n}{b+1}\right\}$$

есть дробная часть числа

$$\frac{1}{b}\log_q\binom{n}{b+1}.$$

Замечание 1. В отличие от (5) оценки (8) охватывают и случай $b\to\infty$. Кроме того, при фиксированном b они справедливы для более широкого диапазона значений

$$m=\left[\frac{1}{b}\log_q\binom{n}{b+1}\right]+k.$$

Остановимся на наиболее интересном частном случае, когда значения параметров b и k фиксированы.

Следствие 3. *Пусть $n\to\infty$, а параметры b и k постоянны. Тогда*

$$\mathsf{P}\left\{\varkappa_q(n,b)\leqslant\frac{1}{b}\log_q\binom{n}{b+1}+k\right\}=\exp\left\{-q^{-b(k-l(n,b))}\right\}\left(1+O(n^{-1/b})\right) \quad (13)$$

и

$$\mathsf{P}\left\{\varkappa_q(n,b)=\left[\frac{1}{b}\log_q\binom{n}{b+1}\right]+k\right\} = \exp\left\{-q^{-b(k-l(n,b))}\right\}\left(1-\exp\left\{-q^{-b(k-l(n,b))}\right\}\right)\left(1+O(n^{-1/b})\right). \quad (14)$$

Замечание 2. Равенство (13) дает лучшую по порядку оценку остаточного члена, чем равенство (5). При этом надо иметь в виду, что выражения в показателе экспоненты в главных членах (5) и (13) несколько отличаются друг от друга.

Замечание 3. Благодаря формулам

$$\begin{aligned}\mathsf{E}\varkappa_q(n,b)&=1+\sum_{m=1}^{\infty}(1-F_{q,n,b}(m)),\\ \mathsf{E}q^{\varkappa_q(n,b)}&=1+(q-1)\sum_{m=1}^{\infty}q^m(1-F_{q,n,b}(m)),\end{aligned} \quad (15)$$

неравенства (8) могут быть использованы для оценивания указанных средних значений. Для $q=2$ соответствующие асимптотические формулы были найдены в [2]. Здесь мы отметим теорему 3 из [2], показывающую, в частности, что при $n\to\infty$ и $b\geqslant 2$

$$\mathsf{E}2^{\varkappa_2(n,b)}\sim n^{1+1/b}. \quad (16)$$

Мы не будем касаться вопроса о явных оценках для средних из-за громоздкости их вывода. Отметим лишь относительно простое утверждение для случая $b = 1$.

Следствие 4. *При любом* $q \geqslant 2$

$$\mathbf{E} q^{\varkappa_q(n,1)} = \infty. \tag{17}$$

Этот результат, по-видимому, не является новым. Однако о нем не упоминается ни в книге [1], ни в статье [2]. Мы обращаем на него внимание по той причине, что он объясняет, почему цифровой поиск всегда редуцируется к случаю b-боров с $b \geqslant 2$ или обрывается каким либо иным способом (с переходом к иным методам, например, к последовательному поиску). Этот вопрос, с несколько иных позиций, обсуждается также в [1, стр. 577].

Доказательство теоремы 1. Пути q-арного b-бора, построенного по n ключам, могут быть отождествлены естественным образом с q-ичными правильными дробями и, тем самым, с точками полуинтервала $[0,1)$. При этом любые точки, попавшие в полуинтервал

$$[jq^{-t}, (j+1)q^{-t}), \qquad j = 0, 1, \ldots, q^t - 1,$$

отвечают путям по бору, имеющим по крайней мере t одинаковых первых звеньев. Поэтому, отождествив процедуру хеширования n слов с равномерным разбрасыванием n точек по полуинтервалу $[0,1)$, мы сопоставим событию $\{\varkappa_q(n,b) \leqslant m\}$ событие, заключающееся в том, что ни один из полуинтервалов

$$[jq^{-m}, (j+1)q^{-m}), \qquad j = 0, 1, \ldots, q^m - 1,$$

не содержит более, чем b точек. Переходя к терминологии теории случайных размещений, получаем следующее.

Пусть n частиц размещаются независимо и с равными вероятностями по $N = q^m$ ячейкам. Обозначим X_i номер ячейки, в которую попала i-ая частица, $i = 1, \ldots, n$. Случайные величины $X_1, \ldots, X_n$ независимы и одинаково распределены, причем

$$\mathbf{P}\{X_1 = k\} = q^{-m} = N^{-1}, \qquad k = 1, 2, \ldots, N.$$

Введем случайную величину

$$\xi_r(n,m) = \sum_{1 \leqslant i_1 < i_2 < \ldots < i_r \leqslant n} I\{X_{i_1} = X_{i_1} = \ldots = X_{ir}\},$$

где $I\{A\}$ — индикатор события A. Приведенные выше рассуждения означают, что

$$\{\varkappa_q(n,b) \leqslant m\} = \{\xi_{b+1}(n,N) = 0\},$$

и, следовательно,

$$\mathbf{P}\{\varkappa_q(n,b) \leqslant m\} = \mathbf{P}\{\xi_{b+1}(n,N) = 0\}. \tag{18}$$

Это тождество позволяет привлечь результаты теории случайных размещений. Введем обозначения

$$\lambda_r = \binom{n}{r} N^{1-r}, \qquad \alpha_r = r^2 \lambda_r n^{-1}(2\lambda_r + r + 3).$$

Лемма 1. *Пусть* $\alpha_2 \leqslant e^{-1}$. *Тогда*

$$\begin{aligned} 1 - \lambda_3 - (1+2\alpha_2)(1+nN^{-1})4\lambda_2^2 n^{-1} \\ < e^{\lambda_2}\mathbf{P}\{\xi_2(n,N)=0\} \\ < (1-(1+2\alpha_2)(1+2nN^{-1})6\lambda_3 - 2\lambda_2^2 n^{-1})_+^{-1}. \end{aligned}$$

Лемма 2. *Пусть* $n > r \geqslant 3$, $rnN^{-1} \leqslant 1$, *а* $\alpha_r \leqslant e^{-1}$. *Тогда*

$$\begin{aligned} 1 - \lambda_{r+1} - (1+3\alpha_r)r^2\lambda_r^2 n^{-1} < e^{\lambda_r}\mathbf{P}\{\xi_r(n,N)=0\} \\ \leqslant \left(1-(1+2\alpha_r)r(r+1)\lambda_{r+1}e^{rn/2N} - \frac{r^2\lambda_r^2}{2n}\right)_+^{-1}. \end{aligned}$$

Доказательства этих утверждений, причем не только для равновероятной, а для произвольной схемы размещения частиц по ячейкам, содержатся как промежуточные этапы рассуждений в доказательствах теорем 1 и 2 работы [5]. По этой причине мы их здесь не приводим. Осталось заметить, что утверждение теоремы 1 следует из сформулированных лемм и тождества (15). Теорема 1 доказана.

Следствие 1 вытекает, после очевидных преобразований из (8) и равенства

$$\mathbf{P}\{\varkappa_q(n,b)=m\} = \mathbf{P}\{\varkappa_q(n,b)\leqslant m\} - \mathbf{P}\{\varkappa_q(n,b)\leqslant m-1\}.$$

Доказательство равенства (2). Воспользуемся указанной выше аналогией с задачей о размещении частиц. Пусть $\mu_r(n,N)$ — число ячеек, содержащих по r частиц каждая после завершения процесса размещения n частиц по N ячейкам. Очевидно, что равенство (18) можно дополнить соотношением

$$\{\xi_{b+1}(n,N)=0\} = \{\mu_r(n,N)=0, r=b+1,\dots,n\}. \tag{19}$$

Комбинируя (18) и (19) с формулой (2) из §1, гл. 11 монографии [6], мы приходим к равенству (2).

Доказательство следствия 2. Прежде всего заметим, что из условий (11) следуют неравенства

$$n > 2e(b+1)^2(b+4)^2, \tag{20}$$

$$k \geqslant 3 - 2b^{-1}\log_q n, \tag{21}$$

$$1 \geqslant (b+1)nq^{-m}, \tag{22}$$

где

$$m = \left[\frac{1}{b}\log_q\binom{n}{b+1}\right] + k.$$

Далее мы хотим воспользоваться результатом теоремы 1, в связи с чем необходимо проверить, что условия следствия 2 обеспечивают выполнимость предположений теоремы 1. Ввиду (22) в проверке нуждается лишь условие (6), которое мы запишем, с учетом определения величины $\alpha_b(n,m)$ в виде

$$(b+1)^2 n^{-1}\gamma(2\gamma+b+4) \leqslant e^{-1}, \tag{23}$$

где, как и прежде,

$$\gamma = \binom{n}{b+1} q^{-bm} = q^{-b(k-l(n,b))}.$$

Сделаем это неравенство более жестким, пусть

$$e(b+1)^2 n^{-1}(2\gamma + b + 4)^2 \leqslant 2.$$

С учетом положительности γ последнее соотношение эквивалентно неравенству

$$\gamma \leqslant \frac{1}{b+1}\sqrt{\frac{n}{2e}} - \frac{b+4}{2}. \tag{24}$$

Поскольку $\gamma \leqslant q^{b(1-k)}$, неравенство (24) будет доказано, если мы установим справедливость неравенств

$$b + 4 < \frac{1}{b+1}\sqrt{\frac{n}{2e}},$$

$$b(1-k) \leqslant \frac{1}{2}(\log_q n - \log_q((b+1)\sqrt{2e})).$$

Но первое из этих неравенств эквивалентно (20), в то время как второе следует из (21) ввиду оценки

$$b^{-1}\log_q(2\sqrt{2e}(b+1)) \leqslant 2.$$

Этого достаточно для доказательства (24) и, таким образом, (23) справедливо, все условия теоремы 1 выполнены, и мы можем использовать соотношение (8).

Из оценок (8) легко следует, что левая часть (12) не превосходит величины

$$\Delta = \max\{1 - A_b(n,m), B_b(n,m) - 1\}. \tag{25}$$

Оценим каждое из выражений, стоящих под знаком максимума в (25) отдельно.

Поскольку $C_b(n,m) = 2(b+1)^2 n^{-1}\gamma^2$ и, в силу (23), $\gamma \leqslant n$, то

$$C_b(n,m) \leqslant 2(b+1)^2 n^{-1/b}\gamma^{1+1/b}. \tag{26}$$

Кроме того, используя соотношение

$$\binom{n}{k} \geqslant \binom{n}{k}^k, \qquad n \geqslant k \geqslant 1,$$

нетрудно убедиться в справедливости следующей цепочки оценок:

$$\binom{n}{b+2} q^{-(b+1)m} \leqslant \frac{n}{b+1} q^{-m}\gamma = \frac{n}{(b+1)\binom{n}{b+1}^{1/b}}\gamma^{1+1/b}$$

$$\leqslant \frac{n}{b+1}\frac{(b+1)^{1+1/b}}{n^{1+1/b}}\gamma^{1+1/b} = (b+1)^{1/b} n^{-1/b}\gamma^{1+1/b}. \tag{27}$$

Комбинируя (26), (27) и (6), получим, что

$$1 - A_b(n,m) \leqslant \binom{n}{b+2} q^{-(b+1)m} + 5C_b(n,m) \leqslant 11(b+1)^2 n^{-1/b}\gamma^{1+1/b}. \tag{28}$$

Далее, вспоминая (22), выводим, что

$$D_b(n,m) \leqslant (b+1)(b+2)\binom{n}{b+1}^{-1}\binom{n}{b+2}q^{-m}\gamma\sqrt{e} \leqslant (b+1)nq^{-m}\gamma\sqrt{e}$$
$$\leqslant (b+1)n\binom{n}{b+1}^{-1/b}\gamma^{1+1/b}\sqrt{e} \leqslant n^{-1/b}\gamma^{1+1/b}(b+1)^{2+1/b}\sqrt{e}. \tag{29}$$

В силу условия (6)

$$B_b(n,m) \leqslant (1 - 2D_b(n,m) - C_b(n,m))_+^{-1},$$

поэтому при

$$2D_b(n,m) + C_b(n,m) \leqslant 1/2 \tag{30}$$

справедливо неравенство

$$B_b(n,m) - 1 \leqslant 2(2D_b(n,m) + C_b(n,m)). \tag{31}$$

Покажем, что в условиях следствия 2 соотношение (30) всегда выполняется. Принимая во внимание оценки (26) и (27), мы видим, что для доказательства (30) достаточно убедиться в том, что

$$2(\sqrt{e}+1)(b+1)^{2+1/b}n^{-1/b}\gamma^{1+1/b} \leqslant 1$$

или, ввиду оценки $\gamma \leqslant q^{b(1-k)}$, что

$$2(\sqrt{e}+1)(b+1)^{2+1/b}n^{-1/b}q^{(b+1)(1-k)} \leqslant 1.$$

Откуда, после логарифмирования и элементарных преобразований получаем, что (30) справедливо, если

$$k \geqslant 1 - \frac{1}{b(b+1)}\log_q n + \frac{(2b+1)}{b(b+1)}\log_q(b+1) + \frac{1}{b+1}\log_q(2\sqrt{e}+1). \tag{32}$$

Нетрудно показать, что сумма двух последних слагаемых в правой части (32) не превосходит 3. Поэтому (32) и (30) следуют из условия (12).

Теперь мы можем воспользоваться (31) и получить оценку

$$B_b(n,m) - 1 \leqslant 2(\sqrt{e}+1)(b+1)^{2+1/b}n^{-1/b}\gamma^{1+1/b}.$$

Объединяя это неравенство с (28) и (25), приходим к (12).

Следствие 2 доказано.

Вывод следствия 3 из следствия 2 аналогичен выводу следствия 1 из теоремы 1, и, в связи с этим, мы опускаем соответствующие рассуждения.

Доказательство следствия 4. Нетрудно проверить, что для всех достаточно больших значений m найдется величина $C = C(n) < \infty$ такая, что

$$B_1(n,m) \leqslant 1 + Cq^{-2m}$$

и $\alpha_1(n,m) \leqslant e^{-1}$. Для таких значений m мы можем использовать теорему 1 при $b = 1$ и, в частности, получить, что

$$\begin{aligned} 1 - F_{q,n,1}(m) &\geqslant 1 - B_1(n,m)\exp\left\{-\binom{n}{2}q^{-m}\right\} \\ &\geqslant 1 - (1 + Cq^{-2m})\left(1 - \frac{1}{2}\binom{n}{2}q^{-m}\right) \\ &\geqslant \frac{1}{2}\binom{n}{2}q^{-m} - C(n)q^{-2m}, \qquad m \geqslant m_0. \end{aligned}$$

Поэтому при всех $m \geqslant m_0$

$$1 - F_{q,n,1}(m) \geqslant C_1(n)q^{-2m}$$

для некоторой константы $C_1(n) > 0$. Подставив эту оценку во второе из равенств (15), получим (17). Следствие 4 доказано.

Список литературы

1. Кнут Д. *Искусство программирования для ЭВМ.* Мир, Москва, 1978.
2. Flajolet P. On the performance evaluation of extendible hashing and trie searching. *Acta Informatica* (1983) **20**, 345–369.
3. Régnier M. On the average height of trees in digital searching and dynamic hashing. *Inform. Proc. Lett.* (1981) **13**, 64–66.
4. Yao A. A note on the analysis of extendable hashing. *Inform. Proc. Lett.* (1980) **11**, 84–86.
5. Зубков А. М., Михайлов В. Г. Оценка точности пуассоновской аппроксимации в задаче о размещении частиц по ячейкам. *Теория вероятностей и ее применения* (1978) **23**, №4, 819–823.
6. Колчин В. Ф., Севастьянов Б. А., Чистяков В. П. *Случайные размещения.* Наука, Москва, 1976.

Статья поступила 02.12.93.