PUC

Series: Monografias em Ciência da Computação
Nº 7/78

OPEN ADDRESSING HASHING WITH UNEQUAL-PROBABILITY KEYS

by

Gaston H. Gonnet

Departamento de Informática

PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO DE JANEIRO
RUA MARQUÊS DE SÃO VICENTE, 225 – CEP-22453
RIO DE JANEIRO – BRASIL

PUC/RJ - DEPARTAMENTO DE INFORMÁTICA

Series: Monografias em Ciência da Computação
Nº 7/78

Series Editor: Michael F. Challis March, 1978

OPEN ADDRESSING HASHING WITH UNEQUAL-PROBABILITY KEYS*

by

Gaston H. Gonnet

* This work has been partially sponsored by FINEP

For copies contact:

Rosane T.L.Castilho
Head, Setor de Documentação e Informação
Depto. de Informática - PUC/RJ
Rua Marques de São Vicente, 209 - Gávea
22453 - Rio de Janeiro - RJ - BRasil

INTRODUCTION

The analysis of hashing algorithms is usually done under the assumption that all keys are equally likely to be accessed. In practice, this is generally not true and as Knuth (73) points out, since most probable keys tend to appear first during table creation, our estimates are pessimistic. In this communication we compute the resulting expected number of accesses in a model of open-addressing hashing for several key distributions. The results presented here correspond to tables that were created inserting keys with decreasing probability, i.e. the optimal ordering when we do not know the specific hashing function.

RESULTS

Let $p_1, p_2, \ldots, p_n$ be the probabilities of accessing the keys $k_1, k_2, \ldots, k_n$. We require that $p_1 \geq p_2 \geq \ldots \geq p_n > 0$ and furthermore that the table is created by inserting the keys in decreasing probability order. Let m be the size of the table, $m \geq n$.

We will analyse the open-addressing scheme. Under this scheme collisions are resolved by computing additional hashing functions until an empty position is found. New probe positions for collision resolution are assumed to be independent of the previous ones, or what is called "uniform probing" (Knuth 73). Double hashing or methods without secondary or higher clustering (Knuth 73, Guibas & Szemeredi 76, Guibas 76) have a performance similar to that of uniform probing.

Under these conditions the i^{th} inserted key requires an average of (Knuth 73)

$$E(\text{accesses for the } i^{th} \text{ key}) = \frac{(m+1)}{(m+2-i)}$$

Consequently the expected number of accesses is

$$E(accesses) = (m+1) \sum_{i=1}^{n} \frac{p_i}{m+2-i}$$

The following results are straightforward to derive, except for the exponential distributing where we have to consider 3 different cases, and the 80%-20% rule. The assymptotic expansion results assume that $m \to \infty$ and that $\alpha = n/m$, $0 < \alpha < 1$, is a constant independent of m.

The complete derivation of the asymptotic results, and the explicit expression of some constants together with some properties of the distributions, can be found in Gonnet (78).

(TABLE I)

.In the above tables of results the asymptotic expansions (last two columns), unless explicitly stated require an $O(ln(m)/m)$ extra term.

We denote by H_n the n^{th} harmonic number (Knuth 68), i.e.

TYPE OF DISTRIBUTION	P_i	CLOSED FORM	$\alpha = n/m$	FULL TABLE
Uniform	$\frac{1}{n}$	$\frac{(m+1)}{n}(H_{m+1} - H_{m-n+1})$	$-\alpha^{-1}\ ln(1-\alpha)$	$ln(m) - 1 + \gamma$
Wedge (general)	$\frac{2(b-i)}{n(2b-n-1)}$	$\frac{2(m+1)}{n(2b-n-1)}$ $\times [(n + (b-m-2)(H_{m+1} - H_{m-n+1})]$		
Wedge ($b = n+1$)	$\frac{2(n+1-i)}{n(n+1)}$		$2\alpha^{-2}(\alpha + (1-\alpha)\ ln(1-\alpha))$	$2 - O(ln(m)/m)$
Zipf's law (harmonic)	$\frac{1}{iH_n}$	$\frac{(m+1)}{(m+2)}\left[1 + \frac{H_{m+1} - H_{m-n+1}}{H_n}\right]$	$1 - \frac{ln(1-\alpha)}{ln(n) + \gamma}$	$2 - \frac{1}{ln(m) + \gamma}$
Bi-harmonic	$\frac{1}{i^2 H_n^{(2)}}$	$\frac{(m+1)}{H_n^{(2)}}\left[\frac{H_n^{(2)}}{m+2} + \frac{H_{m+1} - H_{m-n+1} + H_n}{(m+2)^2}\right]$	$1 + \frac{6(ln(n) + \gamma - ln(1-\alpha))}{m\pi^2}$	$1 + O(ln(m)/m)$
Exponential $(1-a)m > O(1)$	$\frac{(1-a)a^{i-1}}{1-a^n}$		$(1-a^n)^{-1}(1 + \frac{a}{(1-a)m})$	same
$(1-a)m = O(1)$	"		$\frac{(1-a)a^{m+2}}{a-a^{n+1}}[Ei(-ln(a)(m+1))$ $- Ei(-ln(a)(m-n+1))]$	same
$(1-a)m = o(1)$	"		$\frac{(1-a)(m+1)}{1-a^n}[(H_{m+1} - H_{m-n+1})$ $\times (1 - (m+2)(1-a)) - n(1-a)]$	same
80% - 20% rule	$\frac{i^\theta - (i-1)^\theta}{n^\theta}$		$C(\alpha) + O(1/m)$	$1 + \theta\ ln(m) + C_1$

TABLE I

$$H_n = \sum_{i=1}^{n} 1/i = \ln(n) + \gamma + 1/2n + O(n^{-2}) .$$

Similarly

$$H_n^{(2)} = \sum_{i=1}^{n} i^{-2} = \frac{\pi^2}{6} - 1/n + \frac{1}{2n^2} + O(n^{-3}) ,$$

where γ is the Euler's constant, $\gamma = 0.57721\ 56649...$, $Ei(x)$ is the exponential integral (Abramowitz & Stegun eq. 5.1.2, 64)

$$Ei(x) = \int_{-x}^{\infty} e^{-t}\, t^{-1}\, dt ,$$

$$\theta = \ln(.8)/\ln(.2) = 0.13864\ 6883\ ...$$
$$C_1 = 0.08738\ 7674\ ...$$
$$C(0.5) = 1.0861\ 7377\ ...$$
$$C(0.8) = 1.2046\ 3319\ ...$$
$$C(0.9) = 1.2967\ 4512\ ...$$
$$C(0.95) = 1.3903\ 1666\ ...$$

The following table compares some exact numerical results for two tables sizes (100 and 1000) and several occupation factors, for the different distributions, rounded to 5 decimal places

Distribution	50%	80%	90%	95%	100%
uniform	1.37050 1.38468	1.95930 2.00633	2.44353 2.54609	2.92079 3.12697	4.23925 6.49296
wedge (b=n+1)	1.21978 1.22664	1.47789 1.49245	1.62904 1.65115	1.73907 1.76966	1.91605 1.98703
harmonic	1.13951 1.10072	1.29967 1.21957	1.41440 1.30887	1.51982 1.39779	1.79140 1.86468
$1/i^2$	1.02113 1.00354	1.02895 1.00438	1.03333 1.00487	1.03702 1.00531	1.04592 1.00748
exponential (a = 3/4)	1.03201 1.00302	1.03201 1.00302	1.03201 1.00302	1.03201 1.00302	1.03201 1.00302
(a = 9/10)	1.10684 1.00917	1.11327 1.00917	1.11396 1.00917	1.11428 1.00917	1.11477 1.00917
(a = 0.99)	1.33246 1.11936	1.77291 1.12792	2.09726 1.12902	2.40001 1.12958	3.19530 1.13130
80%-20% rule	1.08120 1.08561	1.19327 1.20340	1.27574 1.29445	1.35200 1.38596	1.54897 1.86998

Table II

DERIVATION OF RESULTS

For the uniform distribution we derive:

$$E(\text{number of accesses}) = (m+1)\sum_{i=1}^{n}\frac{p_i}{m+2-i}$$

$$= \frac{(m+1)}{n}\,(H_{m+1} - H_{m-n+1})$$

For n = m, i.e. full table:

$$= \frac{m+1}{m}\,[H_{m+1} - 1] = ln(m) + \gamma - 1 + O(ln(m)/m)$$

and for $\alpha = n/m$

$$= -\alpha^{-1}\, ln(1-\alpha) + O(ln(m)/m)$$

The "wedge" probability distribution is defined by

$$p_i = \frac{2(b-i)}{n(2b-n-1)} \qquad with\ b > n\,.$$

We derive then:

$$E(\#\ of\ accesses) = \frac{2(m+1)}{n(2b-n-1)}\sum_{i=1}^{n}\frac{b-i}{m+2-i}\,,$$

$$= \frac{2(m+1)}{n(2b-n-1)}\,[n + (b-m-2)(H_{m+1} - H_{m-n+1})]\,.$$

For the "typical" wedge, $b = n + 1$ the expected value becomes

$$E(accesses) = 2\alpha^{-2}[\alpha + (1-\alpha)\, ln(1-\alpha)] + O(ln(m)/m)$$

and for a full table, $b - 1 = n = m$:

$$E(accesses) = \frac{2}{m}\,[m - (H_{m+1} - 1)]$$

$$= 2 - O(ln(m)/m) \ .$$

For keys distributed like Zipf's law (or harmonic) we have $p_i = 1/H_n i$. Consequently

$$E(accesses) = \frac{m+1}{H_n} \sum_{i=1}^{n} \frac{1}{(m+2-i)i}$$

$$= \frac{(m+1)}{H_n(m+2)}\,[H_{m+1} - H_{m-n+1} + H_n]$$

$$= \frac{m+1}{m+2}\left[1 + \frac{H_{m+1} - H_{m-n+1}}{H_n}\right]$$

for $n = m$

$$E(accesses) = 2 + O(1/ln(m))$$

and for $\alpha = n/m$

$$E(accesses) = 1 - \frac{ln(1-\alpha)}{ln(n) + \gamma} + O(1/m)$$

For the generalized harmonic we find the similar result

$$E(accesses) = \frac{(m+1)}{(m+a+2)}\left[1 + \frac{H_{m+1} - H_{m-n+1}}{\psi_{(a+n+1)} - \psi_{(a+1)}}\right]$$

when

$$p_i = \frac{1}{i+a} \cdot \frac{1}{[\psi_{(a+n+1)} - \psi_{(a+1)}]} \ .$$

For the bi-harmonic distribution we obtain

$$p_i = \frac{1}{H_n^{(2)} i^2} \ ,$$

$$E(accesses) = \frac{(m+1)}{H_n^{(2)}} \sum_{i=1}^{n} \frac{1}{(m+2-i)i^2}$$

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

$$= \frac{(m+1)}{(m+2)} \left[1 + \frac{H_{m+1} - H_{m-n+1} + H_n}{H_n^{(2)}(m+2)} \right]$$

For $n = m$ we conclude that:

$$E(accesses) = 1 + O(ln(m)/m) \quad ,$$

and for $\alpha = n/m$

$$E(accesses) = 1 + \frac{6(ln\ m + \gamma + ln(\frac{\alpha}{1-\alpha}))}{m\pi^2} + O(m^{-1})$$

In the case that the keys are exponentially distributed we have to consider three different cases to obtain the proper asymptotic expansion.

In all cases we have

$$p_i = \frac{(1-a)a^{i-1}}{1-a^n} \qquad 0 < a < 1$$

a) $m\ ln(a) > O(1)$; that is, for increasing m, $m\ ln(a)$ does not remain bounded.

$$E(accesses) = \frac{(1-a)(m+1)}{1-a^n} \sum_{i=1}^{n} \frac{a^{i-1}}{m+2-i}$$

$$= \frac{(1-a)(m+1)}{1-a^n} \left[\frac{1}{(1-a)(m+1)} + \sum_{i=1}^{n} \frac{(i-1)a^{i-1}}{(m+2-i)(m+1)} \right] ,$$

$$= \frac{(1-a)(m+1)}{1-a^n} \left[\frac{1}{(1-a)(m+1)} + \frac{a}{(1-a)^2(m+1)m} + O((1-a)^{-3}m^{-3}) \right]$$

$$= \frac{1}{1-a^n}\left[1 + \frac{a}{(1-a)m} + O((1-a)^{-2}m^{-2})\right].$$

b) $m\,ln(a) = O(1)$ or that, for increasing m, $m\,ln(a)$ remains bounded.

$$E(accesses) = \frac{(m+1)(1-a)}{1-a^n}\sum_{i=1}^{n}\frac{a^{i-1}}{m+2-i},$$

$$= \frac{(m+1)(1-a)}{1-a^n}\,a^{m+1}\sum_{j=m+2-n}^{m+1}\frac{(1/a)^j}{j}.$$

Expanding $(1/a)^j = e^{-ln(a)j}$ we obtain

$$E(accesses) = \frac{(m+1)(1-a)}{1-a^n}\,a^{m+1}\left[\sum_{j=m+2-n}^{m+1}\frac{1}{j} - \frac{ln(a)j}{j} + \frac{ln^2(a)j^2}{2j} - \dots\right],$$

$$= \frac{(m+1)(1-a)}{1-a^n}\,a^{m+1}\,[E_i(-ln(a)(m+1)) - E_i(-ln(a)(m-n+1)) +$$

$$+ \left.\frac{a^{-x}-1+x\,ln\,a}{2x}\right|_{x=m-n+1}^{x=m+1} + O(n^{-2})]$$

Finally for $m\,ln\,a = o(1)$, we take the derivative of the internal sum with respect to a, for $a = 1$ and use a Taylor expansion around this point i.e.

$$\lim_{a\to 1}\left(\sum_{i=1}^{n}\frac{a^{i-1}}{m+2-i}\right)'_a = \lim_{a\to 1}\sum_{i=1}^{n}\frac{(i-1)a^{i-2}}{m+2-i} = \sum_{i=1}^{n}\frac{i-1}{m+2-i}$$

$$= -n + (m+1)(H_{m+1} - H_{m-n+1})$$

Consequently

$$E(accesses) = \frac{(m+1)(1-a)}{1-a^n}\,[(H_{m+1} - H_{m-n+1})(1-(m+1)(1-a)) - (1-a)n + O(m^2(1-a)^2)]$$

For keys that obey the 80% - 20% distribution we have the following derivation, (the 80% - 20% rule states that the 20% most probable elements add 80% of the total probability and so on recursively, Knuth 73).

We have that

$$p_i = \frac{i^\theta - (i-1)^\theta}{n^\theta},$$

where

$$\theta = \frac{\log(0.8)}{\log(0.2)} = 0.13864\ 68838\ 53213\ 89865\ 9786\ldots$$

The moments of this distribution, computed using the Euler-MacLaurin formula, are given by

$$\mu_1' = \sum_{i=1}^{n} i p_i = \frac{\theta n}{\theta + 1} + \frac{1}{2} - \frac{\zeta(-\theta)}{n^\theta} - \frac{\theta}{12n} + O(n^{-3}) ,$$

$$\mu_2' = \sum_{i=1}^{n} i^2 p_i = \frac{\theta n^2}{\theta + 2} + \frac{\theta n}{\theta + 1} + \frac{2-\theta}{6} - \frac{[2\zeta(-\theta-1) + \zeta(-\theta)]}{n^\theta} + O(n^{-1}) ,$$

$$\mu_3' = \sum_{i=1}^{n} i^3 p_i = \frac{\theta n^3}{\theta + 3} + \frac{3\theta n^2}{2(\theta + 2)} + \frac{\theta(3-\theta)n}{4(\theta + 1)} + O(1) ,$$

$$\mu_k' = \sum_{i=1}^{n} i^k p_i = \frac{\theta n^k}{\theta + k} + \frac{\theta k n^{k-1}}{2(\theta + k - 1)} + \frac{\theta(k-\theta)k n^{k-2}}{12(\theta + k - 2)} + O(n^{k-3}) + O(n^{-\theta}) ,$$

$$\sigma^2 = \frac{\theta n^2}{(\theta + 1)^2(\theta + 2)} + O(n^{1-\theta})$$

But the expected number of accesses can be expressed as

$$E(accesses) = \frac{(m+1)}{(m+2)} \sum_{i=1}^{n} \frac{p_i}{1 - \frac{i}{m+2}}$$

$$= \frac{m+1}{m+2}\left[1 + \sum_{k=1}^{\infty} \frac{\mu_k'}{(m+2)^k}\right]$$

Consequently for $\alpha = n/m \quad 0 < \alpha < 1$,

$$E(accesses) = C(\alpha) + O(m^{-1}) = 1 + \theta \sum_{k=1}^{\infty} \frac{\alpha^k}{\theta + k} + O(m^{-1}) .$$

The last series is not convenient for α near 1, so we use the transformation:

$$\sum_{k=1}^{\infty} \frac{a^k}{\theta+k} = a^{-\theta} \sum_{k=1}^{\infty} \int_0^a x^{k-1+\theta}\, dx = a^{-\theta} \int_0^a x^{\theta} \sum_{k=1}^{\infty} x^{k-1}\, dx$$

$$= a^{-\theta} \int_0^{a} \frac{x^{\theta}}{1-x}\, dx = a^{-\theta} \int_{1-a}^{1} \frac{(1-y)^{\theta}}{y}\, dy$$

$$= a^{-\theta} \int_{1-a}^{1} \left[\frac{1}{y} - \frac{\theta y}{y} + \frac{\theta(\theta-1)}{2y} y^2 - \frac{\theta(\theta-1)(\theta-2)}{3!\, y} y^3 + \dots\right] dy$$

$$= a^{-\theta} \left[\ln y - \theta y + \frac{\theta(\theta-1)}{2!\, 2} y^2 - \frac{\theta(\theta-1)(\theta-2)}{3!\ \ 3} y^3 + \dots \right]_{1-a}^{1}$$

Since

$$- \theta + \frac{\theta(\theta-1)}{2!2} - \frac{\theta(\theta-1)(\theta-2)}{3!\ 3} + \dots =$$

$$= \lim_{b \to 0} \left[\frac{(-\theta)b}{1!\ 1!} + \frac{(-\theta)(1-\theta)b(b+1)}{2!\ 2!} + \frac{(-\theta)(1-\theta)(2-\theta)b(b+1)(b+2)}{3!\ 3!} \dots \right] b^{-1}$$

$$= \lim_{b \to 0} \left[\frac{{}_2F_1(-\theta,b;1;1) - 1}{b} \right] = \lim_{b \to 0} \left[\frac{\Gamma(1)\ \Gamma(1+\theta-b)}{[\Gamma(1+\theta)\ \Gamma(1-b)]} - 1 \right] b^{-1}$$

$$= - \psi(1 + \theta) + \psi(1) = - \psi(1 + \theta) - \gamma$$

we derive the final expression:

$$\sum_{k=1}^{\infty} \frac{a^k}{\theta+k} = a^{-\theta}\left[-\psi(1+\theta) - \gamma - ln(1-a) + \theta(1-a) - \frac{\theta(\theta-1)(1-a)^2}{4} + \frac{\theta(\theta-1)(\theta-2)(1-a)^2}{18} \cdots\right.$$

and we can express the expected value also as

$$E(accesses) = C(\alpha) + O(m^{-1}) = 1 + \theta\alpha^{-\theta}[-\psi(1+\theta) - \gamma - ln(1-\alpha) + \theta(1-\alpha) - \ldots]$$
$$+ O(m^{-1})$$

Direct computation shows that

$C(0.5) = 1.08617\ 37741\ 28045\ 31512\ 1541\ldots$

$C(0.8) = 1.20463\ 31959\ 45617\ 83466\ 2159\ldots$

$C(0.9) = 1.29674\ 51213\ 95053\ 49342\ 7795\ldots$

$C(0.95) = 1.39031\ 66674\ 21293\ 92038\ 3669\ldots$

$C(0.99) = 1.61076\ 56741\ 38615\ 66525\ 3015\ldots$

$C(0.99940\ 04851\ 43161\ 88579\ 4781\ldots) = 2.0$

$C(0.99999\ 95583\ 01845\ 90154\ 1475\ldots) = 3.0$

For full tables, i.e. $n = m$, using the inequality

$$\frac{\theta}{n^\theta}\int_{i-1}^{i} \frac{x^{\theta-1}}{m+2-x}\,dx \le \frac{i^\theta - (i-1)^\theta}{n^\theta(m+2-i)} \le \frac{\theta}{n^\theta}\int_{i-1}^{i}\frac{x^{\theta-1}}{m+1-x}\,dx$$

then

$$\frac{(m+1)\theta}{n^\theta}\int_0^{n/m+2}\frac{(m+2)^{\theta-1}}{1-y}\,y^{\theta-1}\,dy \le E(accesses) \le \frac{(m+1)\theta}{n^\theta}\int_0^{n/m+1}\frac{(m+1)^{\theta-1}y^{\theta-1}}{1-y}\,dy$$

and using the same derivation as before we find that for full tables

$$\left|E(accesses) - [1+\theta(ln(m) - \frac{ln\ 2}{2} - \psi(1+\theta) - \gamma)]\right| \le \frac{\theta\ ln\ 2}{2}$$

Let

$$E(accesses) = 1 + \theta\ ln(m) + C_1 + O(m^{-1})\ ,$$

then if we use the first five terms of the expansion of μ_k':

$$E(\mathit{accesses}) = \frac{m+1}{m+2}\left[1 + \sum_{k=1}^{\infty} \frac{\theta n^k}{(\theta+k)(m+2)^k} + \sum_{k=1}^{\infty} \frac{\theta n^{k-1}}{2(m+2)^k} + \sum_{k=1}^{\infty} \frac{k\theta n^{k-2}}{12(m+2)^k} - \right.$$

$$\left. + \sum_{k=1}^{\infty} \frac{k\theta(\theta-1)n^{k-3}}{24(m+2)^k} - \sum_{k=1}^{\infty} \frac{k^3\theta n^{k-4}}{720(m+2)^k}\right] + O\left(\frac{k^2 n^{k-4}}{m^k}\right)$$

For $n = m$

a) $$\sum_{k=1}^{\infty} \frac{\theta n^k}{(\theta+k)(m+2)^k} = \theta \ln(m) + C_1 + O(m^{-1})$$

b) $$\sum_{k=1}^{\infty} \frac{\theta n^{k-1}}{2(m+2)^k} = \frac{\theta}{4} + O(m^{-1})$$

c) $$\sum_{k=1}^{\infty} \frac{k\theta n^{k-2}}{12(m+2)^k} = \frac{\theta}{48} + O(m^{-1})$$

d) $$\sum_{k=1}^{\infty} \frac{k\theta(\theta-1)n^{k-3}}{24(m+2)^k} = O(m^{-1})$$

e) $$\sum_{k=1}^{\infty} \frac{k^3\theta n^{k-4}}{720(m+2)^k} = \frac{\theta}{1920} + O(m^{-1})$$

So we can bound C_1 by

$$\theta\left[-\frac{1}{1920} + \frac{1}{48} + \frac{1}{4} - \ln(2) - \gamma - \psi(1+\theta)\right] \le C_1 \le \theta\left[\frac{1}{48} + \frac{1}{4} - \ln(2) - \gamma - \psi(1+\theta)\right]$$

$$-\,0.0873946\ldots \le C_1 \le -\,0.0873224\ldots$$

From an asymptotically equivalent distribution (i.e. $P_i = \alpha i^{\theta-1}$) we derive the strong conjecture

$$C_1 = -\theta(\psi(1+\theta)+1) = -0.08738\ 76749\ 82611\ 29115\ 5901\ldots$$

$$(\psi(1+\theta) = -0.36971\ 05008\ 49560\ 89275\ 0609\ldots)$$

CONCLUSIONS

The access probability distributions analysed show that, if the tables are constructed in the adequate order, the average number of accesses remains very low even for full tables (for many distributions it is always below 2 accesses). In practical applications we usually don't know the probability distributions. On the other hand, elements with higher accessing probability appear first and are inserted first. The real expected values (which is a much harder problem) lie somewhere between the values found and the values for the uniform distribution.

ACKNOWLEDGMENTS

The author wishes to acknowledge the contribution done by Tatuo Nakanishi in the computations of the 80%-20% rule.

REFERENCES

1. Abramowitz, M. and Stegun, I. A. Handbook of Mathematical Functions. Dover Publications, New York, 1964.

2. Gonnet, G. H. Open-addressing hashing with unequal-probability keys. Monografias em Ciência da Computação 7/78, PUC Rio de Janeiro, Brazil.

3. Guibas, L. J. The analysis of hashing algorithms that exhibit K-ary clustering. Proceedings of the 17th annual IEEE-FOCS Symposium, Houston Texas, October 1976, pp 183-196.

4. Guibas, L. J. and Szemeredi, E. The analysis of double hashing. Proceedings of the 8th ACM Symposium on the theory of Computing, Hershey Pennsylvania, May 1976, pp. 187-191.

5. Knuth, D. E. The Art of Computer Programming. Addison-Wesley, Don Mills Ontario, 1968 & 1973, vol. I & III, chap. I & VI.