24

Multidimensional Digital Searching

Helmut Prodinger
Technical University of Vienna

[résumé par Danièle Gardy]

Cet exposé présente les performances de la recherche partiellement spécifiée, sur des clés multidimensionnelles stockées dans des *tries*, ou dans d'autres structures digitales.

1 Présentation

Les *"tries"* (arbres lexicographiques, arbres préfixes) sont habituellement construits à partir de clés qui sont des mots infinis sur l'alphabet $\{0,1\}$; les nœuds internes servent à se diriger dans l'arbre et les clés sont stockées aux feuilles; on ne garde des clés que ce qui suffit à les distinguer des autres clés lors de l'insertion. On va s'intéresser dans cet exposé à leurs performances, lorsqu'on les étend à des clés multidimensionnelles.
Structures m-dimensionnelles : Ici, les clés sont des ensembles de m-uplets : $K = (K_1, K_2, \cdots, K_m)$, où les K_i sont des suites de bits. Pour construire la structure arborescente, on remplace d'abord chaque clé composée K par une chaîne de bits obtenue comme suit : on prend le bit initial de chaque composante K_i (dans l'ordre), puis le deuxième bit, etc. Ensuite, on construit le *trie* sur la chaîne de bits, comme dans le cas unidimensionnel.
Le cas le plus simple est obtenu lorsque $m = 2$; c'est celui qui sera traité en détail dans la suite. Par exemple, pour $m = 2$, la clé $K = (01\ldots, 00\ldots)$ donne la chaîne $0010\ldots$, et la clé $(10\ldots, 01\ldots)$ donne la chaîne $1001\ldots$ Une *recherche partiellement spécifiée* consiste à chercher toutes les clés dont certaines coordonnées sont égales à une valeur spécifiée, les autres coordonnées pouvant prendre n'importe quelle valeur. Le profil (*search pattern*) est un mot de $\{*, S\}^m$: S correspond à une coordonnée spécifiée, et $*$ à une coordonnée non spécifiée. Une mesure de complexité classique pour ce problème est le nombre de nœuds internes visités. Attention, certains nœuds internes peuvent être visités même s'ils conduisent à une feuille contenant une clé qui ne répond pas à la requête.

2 Etude détaillée du cas m=2

Soit le profil $\omega = (*, S)$: on cherche donc une clé dont la première composante est quelconque, et la seconde égale à une valeur donnée. Soit $\omega' = (S, *)$ le profil se déduisant de ω par permutation circulaire.

2.1 Coût moyen d'une recherche suivant le profil ω

Notons $H_n^\omega(z)$ la fonction génératrice de probabilité du coût de la requête de profil ω sur des données de taille n (i.e. le *trie* est construit à partir de n clés bidimensionnelles) :

$$H_n^\omega(z) = \sum_{i \geq 0} \text{Proba}(\omega \text{ visite } i \text{ nœuds internes})\, z^i.$$

On définit de même $H_n^{\omega'}(z)$. Remarquons au passage que $H_n^{\omega}(1) = H_n^{\omega'}(1) = 1$.
Lors d'une recherche partiellement spécifiée dans un *trie*, et si la première coordonnée n'est pas spécifiée, la recherche se poursuit dans les deux sous-arbres; par contre, lorsque la première coordonnée est spécifiée, on ne poursuit la recherche que dans un seul sous-arbre (le gauche ou le droit, suivant la valeur qui est spécifiée). Dans les deux cas, la recherche au niveau inférieur se fait suivant le motif obtenu en décalant d'un bit le motif initial.
Si l'on suppose que les bits 0 et 1 ont même probabilité (hypothèse faite pour toute la suite), la probabilité que k clés parmi les n commencent par un bit donné est $\binom{n}{k}2^{-n}$. On obtient donc les équations de récurrence suivantes, liant le coût de la recherche suivant le motif $\omega = (*, S)$ au coût de la recherche suivant le motif permuté $\omega' = (S, *)$:

$$\begin{aligned} H_n^{\omega}(z) &= z\sum_{k=0}^{n} 2^{-n}\binom{n}{k} H_k^{\omega'}(z)\, H_{n-k}^{\omega'}(z); \\ H_n^{\omega'}(z) &= z\sum_{k=0}^{n} 2^{-n}\binom{n}{k} H_k^{\omega}(z). \end{aligned}$$

De manière générale, si on travaille avec des clés de dimension m, on aura m équations, chacune définissant la fonction génératrice pour un des profils obtenus par rotation du profil initial, à partir des $m-1$ autres fonctions génératrices associées aux profils permutés.
Revenons à l'exemple précédent : les valeurs moyennes $l_n^{\omega} = (H_n^{\omega})'(1)$ et $l_n^{\omega'} = (H_n^{\omega'})'(1)$, pour $n \geq 2$, satisfont les équations de récurrence suivantes, avec les conditions initiales $l_0^{\omega} = l_1^{\omega} = 1$:

$$l_n^{\omega} = 1 + 2\sum_{k=0}^{n} 2^{-n}\binom{n}{k} l_k^{\omega'}, \qquad l_n^{\omega'} = 1 + \sum_{k=0}^{n} 2^{-n}\binom{n}{k} l_k^{\omega}. \tag{1}$$

Définissons les fonctions génératrices exponentielles de l_n^{ω} et de $l_n^{\omega'}$: $L^{\omega}(z) = \sum_{n\geq 0} l_n^{\omega} z^n/n!$ et $L^{\omega'}(z) = \sum_{n\geq 0} l_n^{\omega'} z^n/n!$, puis les fonctions génératrices exponentielles de Poisson $\tilde{L}^{\omega}(z) = e^{-z}L^{\omega}(z)$ $= \sum_{n\geq 0} \tilde{l}_k^{\omega} z^n/n!$ et $\tilde{L}^{\omega'}(z) = e^{-z}L^{\omega'}(z) = \sum_{n\geq 0} \tilde{l}_k^{\omega'} z^n/n!$. Les coefficients de L^{ω} et $\tilde{L}^{\omega}$ sont liés par : $l_n^{\omega} = \sum_{k=0}^{n} \binom{n}{k}\tilde{l}_k^{\omega}$; de même ceux de $L^{\omega'}$ et de $\tilde{L}^{\omega'}$.
Les équations (1) sur les l_n^{ω} se traduisent d'abord en un système d'équations fonctionnelles sur les fonctions génératrices exponentielles L^{ω} et $L^{\omega'}$, puis en un système analogue sur les fonctions génératrices de Poisson :

$$\begin{aligned} \tilde{L}^{\omega}(z) &= 1 - (1+z)e^{-z} + 2\tilde{L}^{\omega'}(z/2); \\ \tilde{L}^{\omega'}(z) &= 1 - (1+z)e^{-z} + \tilde{L}^{\omega}(z/2). \end{aligned}$$

On passe alors à un système sur $\tilde{l}_n^{\omega}$ et $\tilde{l}_n^{\omega'}$, qu'on peut résoudre explicitement. Ainsi, on obtient :

$$\tilde{l}_n^{\omega} = (-1)^n \frac{(n-1)(1+2^{1-n})}{1-2^{1-2n}},$$

ce qui donne la moyenne l_n^{ω} sous forme de somme :

$$l_n^{\omega} = 1 + \sum_{k=1}^{n}\binom{n}{k}(-1)^k f(k), \qquad \text{avec} \qquad f(k) = \frac{(k-1)(1+2^{1-k})}{1-2^{1-2k}}. \tag{2}$$

La fonction f, définie sur les entiers positifs, peut être prolongée en une fonction analytique en posant $f(z) = (z-1)(1+2^{1-z})/(1-2^{1-2z})$. On peut alors utiliser la méthode de Rice, qui sert justement à calculer les sommes de la forme ci-dessus (voir par exemple la présentation qui en est faite dans [2, p. 754]). On a :

$$\sum_{k=1}^{n} \binom{n}{k} (-1)^k f(k) = -\frac{1}{2i\pi} \int_C [N;z] f(z) dz,$$

avec

$$[N;z] = \frac{\Gamma(-z)\Gamma(n+1)}{\Gamma(n+1-z)} = \frac{(-1)^{n+1} n!}{z(z-1)\cdots(z-n)},$$

et où C est une courbe dans le domaine d'analycité de f, entourant les points $1, 2, \cdots n$. La somme (2) vaut asymptotiquement $\sum [N;c] f(c)$, où c décrit l'ensemble des résidus à gauche de la droite $\Re z = 1$ (il suffit d'appliquer la formule des résidus à l'intégrande, qui a pour seuls pôles les entiers positifs). D'après la définition de f, les seuls résidus sont obtenus pour $\Re z = 1/2$, plus précisément pour $z = 1/2(1+\xi_k)$ avec $\xi_k = 2ik\pi/\log 2$. Donc, en posant $L = \log 2$:

$$l_n^\omega \approx \sqrt{n} \left(\sqrt{\pi} \frac{1+\sqrt{2}}{2L} + \tau^\omega(\log_2 \sqrt{n}) \right) \approx \sqrt{n} \left(3,086705281... + \tau^\omega(\log_2 \sqrt{n}) \right), \qquad (3)$$

où τ^ω est une fonction périodique, de moyenne 0 et de période 1, d'amplitude faible, et qui peut s'écrire comme somme d'une série de Fourier :

$$\tau^\omega(x) = \sum_{k \neq 0} \tau_k e^{2ik\pi x} \qquad \text{avec} \qquad \tau_k = \frac{1}{2L}(1+(-1)^k\sqrt{2}) \, \frac{-1+\xi_k}{2} \, \Gamma(\frac{-1-\xi_k}{2}). \qquad (4)$$

On peut transformer l'expression des coefficients τ_k, en utilisant le fait que $\Gamma(z+1) = z\Gamma(z)$; on obtient :

$$\tau_k = \frac{1}{2L}(1+(-1)^k\sqrt{2}) \, \frac{1-\xi_k}{1+\xi_k} \, \Gamma(\frac{1-\xi_k}{2}).$$

2.2 Variance

Le calcul de la variance du coût d'une recherche suivant le profil ω fait intervenir $W_n^\omega = (H_n^\omega)''(1)$. Soient W^ω et $\tilde{W}^\omega$ la fonction génératrice exponentielle et la fonction génératrice de Poisson associées : $W^\omega(z) = \sum_{n \geq 0} W_n^\omega z^n/n!$ et $\tilde{W}^\omega(z) = e^{-z} W^\omega(z)$. On obtient les équations fonctionnelles suivantes :

$$\tilde{W}^\omega(z) = 2\tilde{W}^{\omega'}(z/2) + 4\tilde{L}^{\omega'}(z/2) + 2(\tilde{L}^{\omega'}(z/2))^2, \qquad \tilde{W}^{\omega'}(z) = \tilde{W}^\omega(z/2) + 2\tilde{L}^\omega(z/2).$$

Il est possible d'en déduire une formule explicite, assez compliquée, pour les W_n^ω :

$$W_n^\omega = \sum_{k=2}^{n} \binom{n}{k} (-1)^k f(k),$$

avec

$$\begin{aligned} f(k) \quad &= \quad \frac{(z-1)2^{1-z}}{(1-2^{1-2z})(1-2^{2-2z})} \left(5 + 2^{3-z} + 2^{3-2z} + z2^{z-2} - 2^z + \frac{z}{4} + (z - \frac{9}{2}) \left(\frac{3}{2}\right)^{z-2} \right) \\ &+ \frac{2^{3-2z}}{(1-2^{1-2z})(1-2^{2-2z})} \sum_{l \geq 2} \binom{z}{l} \frac{(l-1)(z-l+1)(1+2^{-l})(1+2^{z-l})2^{-l}}{1-2^{1-2l}}. \end{aligned}$$

On applique là aussi la méthode de Rice; les résidus sont en $1+\xi_k$, et on obtient une expression asymptotique pour W_n^ω, de la forme suivante, pour des coefficients α, β et γ calculables et donnés explicitement dans [5, formule (28)] :

$$W_n^\omega \approx \alpha n + \beta\sqrt{n} + \gamma. \qquad (5)$$

La variance σ^2 vaut $W_n^\omega + l_n^\omega - (l_n^\omega)^2$. Compte tenu des approximations (3) et (5), elle est donc de la forme $\sigma^2 = Bn + A\sqrt{n} + o(\sqrt{n})$. On montre que $B = 0$ (voir la section suivante), et donc que $\sigma^2 \approx A\sqrt{n}$. La variance a alors pour terme principal $A\sqrt{n}$, où, comme pour l'expression asymptotique de la moyenne, A est la somme d'un terme constant A_1 et d'un terme périodique, de moyenne 0 et d'amplitude faible. De plus, on a une forme explicite pour A_1, ce qui permet de le calculer numériquement : $A_1 = 2{,}0918454...$

2.2.1 Preuve que B=0

Le coefficient B est égal à la somme d'un terme périodique et d'un terme constant B_1 :

$$B_1 = (1/L)\left(8 - \frac{7}{12} - 2\sum_{l\geq 2}(-1)^l \frac{(1+2^{-l})(1+2^{1-l})2^{-l}}{1-2^{1-2l}}\right) - \pi\frac{(1+\sqrt{2})^2}{4L^2} - [(\tau^\omega)^2]_0,$$

où $[(\tau^\omega)^2]_0$ est la moyenne du carré de τ^ω; la fonction périodique τ^ω est celle définie dans (4). Quelques manipulations algébriques permettent d'exprimer ce dernier terme :

$$[(\tau^\omega)^2]_0 = \sum_{k\neq 0}\tau_k\tau_{-k} = 2\sum_{k\geq 1}(1+(-1)^k\sqrt{2})^2\,\Gamma(\frac{1-\xi_k}{2})\,\Gamma(\frac{1+\xi_k}{2}).$$

Le produit $\Gamma((1-\xi_k)/2)\Gamma((1+\xi_k)/2)$ peut être simplifié grâce à la formule des compléments $\Gamma(s)\Gamma(1-s) = \pi/\sin(\pi s)$; on montre ainsi qu'il est égal à $\pi/\cos(\pi\xi_k/2)$, ou encore à $2\pi/(e^{k\pi^2/L} + e^{-k\pi^2/L})$. On obtient finalement une expression de $[(\tau^\omega)^2]_0$ faisant intervenir les fonctions $F(x) = \sum_{k\geq 1} e^{-kx}/(1+e^{-2kx})$ et $G(x) = \sum_{k\geq 1}(-1)^{k-1}e^{-kx}/(1+e^{-2kx})$:

$$[(\tau^\omega)^2]_0 = \frac{3\pi}{L^2}\,F\left(\frac{\pi^2}{L}\right) - \frac{2\sqrt{2}\pi}{L^2}G\left(\frac{\pi^2}{L}\right).$$

La fonction F vérifie une équation fonctionnelle (qui sera prouvée dans la section 2.2.2) :

$$F(x) = \frac{\pi}{4x} - \frac{1}{4} + \frac{\pi}{x}F(\frac{\pi^2}{x}) \qquad (0 < x < \pi^2).$$

On obtient donc, en remarquant de plus que $G(x) = F(x) - 2F(2x)$:

$$[(\tau^\omega)^2]_0 = \frac{3-2\sqrt{2}}{L}F(L) + \frac{2\sqrt{2}}{L}F\left(\frac{L}{2}\right) + \frac{3}{4L} - \frac{(3+2\sqrt{2})\pi}{4L^2}.$$

On en tire :

$$B_1 = \frac{20}{3L} + \frac{2\sqrt{2}-3}{L}F(L) - \frac{2\sqrt{2}}{L}F(\frac{L}{2}) - \frac{2}{L}\sum_{n\geq 2}(-1)^n\frac{(1+2^{-n})(1+2^{1-n})2^{-n}}{1-2^{1-2n}}. \qquad (6)$$

Sur cette expression, on peut alors montrer que B_1 est nul (voir section 2.2.3). Donc, B se réduit à un terme périodique, de moyenne nulle. Or ce terme périodique est lui aussi identiquement nul. Supposons en effet qu'il ne le soit pas : multiplié par n, il donnerait alors le terme principal de la variance, qui serait alors négative pour un nombre infini de valeurs de n.

2.2.2 Comment prouver l'équation fonctionnelle sur F?

On développe F :

$$F(x) = \sum_{k\geq 1} \sum_{j\geq 0} (-1)^j e^{-k(2j+1)x}.$$

On a donc une somme harmonique, à laquelle on peut appliquer la transformation de Mellin (voir par exemple [8, p. 453-455] pour la définition et les propriétés de base de la transformée de Mellin en analyse d'algorithmes; on peut aussi regarder, entre autres, [6, p. 18 et suivantes]) :

$$\begin{aligned} F^*(s) &= \int_0^{+\infty} F(x)x^{s-1}dx \;=\; \sum_{k,j}(-1)^j \int_0^{+\infty} e^{-k(2j+1)x}x^{s-1}dx \\ &= \sum_{k,j}(-1)^j k^{-s}(2j+1)^{-s} \int_0^{+\infty} e^{-t}t^{s-1}dt \;=\; \Gamma(s) \sum_{k\geq 1} k^{-s} \sum_{j\geq 0}(-1)^j(2j+1)^{-s}. \end{aligned}$$

Or $\sum_{k\geq 1} k^{-s} = \zeta(s)$; en posant $\beta(s) = \sum_{j\geq 0}(-1)^j(2j+1)^{-s} = 1 - 1/3^s + 1/5^s - \cdots$ on obtient $F^*(s) = \Gamma(s)\,\zeta(s)\,\beta(s)$. Par ailleurs, $F(x) \leq e^{-x}/(1-e^{-x})$, et F a donc comme bande fondamentale $< 1; +\infty >$. La formule d'inversion de Mellin donne alors, en intégrant sur la droite verticale d'équation $\Re(s) = 3/2$ contenue dans cette bande fondamentale :

$$F(x) = \frac{1}{2i\pi} \int_{3/2-i\infty}^{3/2+i\infty} \Gamma(s)\,\zeta(s)\,\beta(s)\,x^{-s}ds.$$

La fonction $\Gamma(s)$ a des pôles simples en $s = 0, -1, -2, \ldots$, et la fonction $\zeta(s)$ a un pôle simple unique en $s = 1$. En ce qui concerne la fonction β, on remarque d'abord qu'elle peut s'exprimer à l'aide de la fonction de Hurwitz $\zeta(a, s) = \sum_{n\geq 0}(n + a)^{-s}$ [9, p. 265] :

$$\beta(s) = \frac{\zeta(s, \frac{1}{4}) - \zeta(s, \frac{3}{4})}{4^s}.$$

Par ailleurs, la fonction $\zeta(s, a)$ est analytique pour tout s, sauf en $s = 1$ qui est un pôle simple [1, p. 255] [9, p. 266]. La fonction β, étant définie au point 1, est donc entière, et $\beta(1) = \pi/4$. Notons aussi que $\beta(0) = 1/2$ et $\zeta(0) = -1/2$. En déplaçant le contour d'intégration à gauche des pôles 0 et 1 (par exemple sur la droite verticale $\Re z = -1/2$), on ajoute les résidus de l'intégrande en ces pôles. Le résidu pour $s = 1$ donne $\Gamma(1)\beta(1)/x = \pi/(4x)$, et celui pour $s = 0$ donne $\beta(0)\zeta(0) = -1/4$. On obtient donc :

$$F(x) = \frac{\pi}{4x} - \frac{1}{4} + \frac{1}{2i\pi} \int_{-1/2-i\infty}^{-1/2+i\infty} \Gamma(s)\,\zeta(s)\,\beta(s)\,x^{-s}ds.$$

On utilise ensuite la formule de duplication de la fonction Γ [7, p. 160] [9, p. 240] :

$$\Gamma(s) = \Gamma(\frac{s}{2})\Gamma(\frac{s+1}{2})\frac{2^{s-1}}{\sqrt{\pi}}, \tag{7}$$

puis [7, p. 160] :

$$\Gamma(\frac{s}{2})\zeta(s) = \pi^{s-1/2}\Gamma(\frac{1-s}{2})\zeta(1-s).$$

La fonction $\zeta(a,s)$ vérifie l'équation fonctionnelle suivante, pour tous h et k entiers tels que $1 \leq h \leq k$: [1, p. 261] :

$$\zeta(1-s,\frac{h}{k}) = \frac{2\Gamma(s)}{(2\pi k)^s}\sum_{r=1}^{k}\cos(\frac{\pi s}{2}-\frac{2\pi rh}{k})\ \zeta(s,\frac{r}{k}).$$

Cette équation, la formule des compléments et la formule de duplication de la fonction Γ permettent d'obtenir l'équation fonctionnelle reliant β et Γ :

$$\beta(1-s)\Gamma(1-\frac{s}{2}) = 2^{2s-1}\pi^{-s+1/2}\Gamma(\frac{s+1}{2})\beta(s).$$

On a alors

$$\Gamma(s)\ \zeta(s)\ \beta(s) = 2^{-s}\pi^{2s-3/2}\ \Gamma(\frac{1-s}{2})\ \Gamma(1-\frac{s}{2})\ \zeta(1-s)\ \beta(1-s),$$

et il suffit d'appliquer la formule (7) à $\Gamma(1-s)$ pour obtenir

$$\Gamma(s)\ \zeta(s)\ \beta(s) = \pi^{2s-1}\ \Gamma(1-s)\ \zeta(1-s)\ \beta(1-s).$$

L'expression intégrale de $F(x)$ devient alors

$$F(x) = \frac{\pi}{4x}-\frac{1}{4}+\frac{1}{2i\pi}\int_{(-1/2)}\pi^{2s-1}\Gamma(1-s)\zeta(1-s)\beta(1-s)x^{-s}ds.$$

Un changement de variable $1-s=u$ permet de conclure :

$$\int_{(-1/2)}\pi^{2s-1}\Gamma(1-s)\zeta(1-s)\beta(1-s)x^{-s}ds = \frac{\pi}{x}\int_{(3/2)}\Gamma(u)\zeta(u)\beta(u)(\frac{x}{\pi^2})^u du,$$

et donc $F(x) = \pi/(4x) - 1/4 + (\pi/x)F(\pi^2/x)$.

On a donc prouvé l'équation fonctionnelle sur F. Ce type d'analyse revient fréquemment dans les analyses de *tries* : pour montrer la nullité d'un coefficient, on est amené à établir des identités fonctionnelles du type $f(x) = <expr> + f(C\pi^2/x)$. A titre d'exemple, voici quelques identités, se prouvant de la même manière que ci-dessus :

$$\begin{aligned}
f(x) &= \textstyle\sum_{k\geq 1}\frac{k}{e^{kx}-1} &&= \frac{\pi^2}{6x^2}-\frac{1}{2x}+\frac{1}{24}-\frac{4\pi^2}{x^2}f(\frac{4\pi^2}{x});\\
g(x) &= \textstyle\sum_{k\geq 1}\frac{(-1)^{k-1}}{k(e^{kx}-1)} &&= \frac{\pi^2}{12x}-\frac{\log 2}{2}+\frac{x}{24}-g(\frac{2\pi^2}{x});\\
h(x) &= \textstyle\sum_{k\geq 1}\frac{1}{k(e^{kx}-1)} &&= \frac{\pi^2}{6x}-\frac{x}{24}+\frac{\log(2\pi x)}{2}+h(\frac{4\pi^2}{x}).
\end{aligned}$$

2.2.3 Nullité de B_1

L'équation (6), légèrement transformée, donne :

$$B_1\ L = \frac{20}{3}+2\sqrt{2}\left(F(L)-F(\frac{L}{2})\right)-3F(L)-2\sum_{n\geq 2}(-1)^n\frac{(1+2^{-n})(1+2^{1-n})2^{-n}}{1-2^{1-2n}},$$

avec $F(x) = \sum_{k\geq 1} e^{-kx}/(1+e^{-2kx})$. On a donc, pour $L = \log 2$:

$$F(L) = \sum_{k\geq 1}\frac{2^{-k}}{1+2^{-2k}};\qquad \text{et}\qquad F(\frac{L}{2}) = \sum_{k\geq 1}\frac{2^{-k/2}}{1+2^{-k}}.$$

On décompose la somme qui intervient dans $F(L/2)$, suivant la parité de k, et on obtient :

$$F(L/2) = F(L) + \sqrt{2}\sum_{p\geq 1}\frac{2^{-p}}{1+2^{1-2p}},$$

ce qui montre que

$$B_1\ L = \frac{20}{3} + 4\sum_{p\geq 1}\frac{2^{-p}}{1+2^{1-2p}} - 3F(L) - 2\sum_{n\geq 2}(-1)^n\frac{(1+2^{-n})(1+2^{1-n})2^{-n}}{1-2^{1-2n}}.$$

Par ailleurs,

$$\begin{aligned} F(L) &= \textstyle\sum_{k\geq 1} 2^{-k}\sum_{j\geq 0}(-1)^j 2^{-2kj} &&= \displaystyle\sum_{j\geq 0}(-1)^j\sum_{k\geq 1}2^{-k(2j+1)} \\ &= \textstyle\sum_{j\geq 0}(-1)^j\frac{2^{-1-2j}}{1-2^{-1-2j}} &&= \displaystyle\sum_{j\geq 1}(-1)^{j+1}\frac{2^{1-2j}}{1-2^{1-2j}}. \end{aligned}$$

Enfin,

$$\frac{2^{-n}(1+2^{-n})(1+2^{1-n})}{1-2^{1-2n}} = -2^{-n} + \frac{2^{1-n}}{1-2^{1-2n}} + 3\frac{2^{-2n}}{1-2^{1-2n}}.$$

En reportant ceci et la dernière expression de F dans B_1, on remarque que la plupart des sommes s'annulent, et on trouve : $B_1\ L = 2/3 + 2\sum_{n\geq 1}(-1)^n 2^{-n} = 0$.

2.3 Profil permuté $\omega' = (S, *)$

Son coût moyen cède au même type d'analyse, et on obtient, avec encore une fois des fonctions τ_1 et τ_2 périodiques, de moyenne 0 et de période 1, et de faible amplitude :

$$\begin{aligned} l_n^{\omega'} &\approx \sqrt{n}\left(\sqrt{\pi}\frac{1+\sqrt{2}}{2\sqrt{2}\log 2} + \tau_1(\log_2\sqrt{n})\right) \approx \sqrt{n}\,(2,182630235\ldots + \tau_1(\log_2\sqrt{n}))\,; \\ \sigma^2 &\approx \sqrt{n}\,(1,15324222\ldots + \tau_2(\log_2\sqrt{n}))\,. \end{aligned}$$

3 Généralisations et extensions

3.1 Autres types de données

La méthode exposée dans la section 2 permet d'analyser les performances de la recherche multidimensionnelle sur d'autres structures : arbres Patricia, arbres de recherche digitaux. Les versions multidimensionnelles de ces structures sont obtenues en construisant comme à l'accoutumée les arbres, sur des clés infinies obtenues à partir des clés composées. Par contre, on n'a que les valeurs moyennes des coûts de recherche, et non les variances.

Notons $l_{\omega,n}^{[A]}$ le nombre moyen de nœuds visités, pour une recherche suivant le profil ω dans un arbre de type A et avec n clés (A vaut D pour un arbre de recherche digital, P pour un arbre Patricia, et T pour un *trie*). On peut alors obtenir le coût moyen d'une recherche partiellement spécifiée : lorsque s attributs parmi m sont spécifiés, le coût est d'ordre $n^{1-s/m}$:

$$l_{\omega,n}^{[A]} \approx C^{[A]} n^{1-s/m}.$$

La constante $C^{[A]}$ diffère selon le type de la structure :

$$
\begin{aligned}
C^{[P]} &= (1+\frac{s}{m})\,(1-2^{-s/m})\,\frac{1}{mL}\,\frac{\Gamma(s/m)}{1-s/m}\,\mathcal{S};\\
C^{[D]} &= \frac{Q(2^{1-s/m})}{Q_\infty}\,\frac{1}{mL}\,\frac{\Gamma(s/m)}{1-s/m}\,\mathcal{S};\\
C^{[T]} &= \frac{s}{m^2L}\,\frac{\Gamma(s/m)}{1-s/m}\,\mathcal{S}.
\end{aligned}
$$

Dans ces formules, $\mathcal{S} = \sum_{j=0}^{m-1}\delta_1\ldots\delta_j 2^{-j(1-s/m)}$, et δ_i vaut 1 si le $i^{ème}$ attribut est spécifié, et 0 sinon. Enfin, la fonction Q est définie par : $Q(x) = (1-x/2)(1-x/4)(1-x/8)\cdots$, et $Q_\infty = Q(1) = 0.28878809...$ Le résultat sur les *tries* avait été obtenu par Flajolet et Puech [4]. On peut comparer les constantes (prendre $x = s/m < 1$, et voir que $(1+x)(1-2^{-x}) < Q(2^{1-x})/Q_\infty < x$) :

$$C^{[P]} < C^{[D]} < C^{[T]}.$$

3.2 Autres types de problèmes

Des récurrences du type de (1) peuvent apparaître dans d'autres problèmes que ceux issus de l'étude d'une structure arborescente, ainsi le dénombrement probabiliste analysé dans [3].

Références

[1] T. M. Apostol. *Introduction to Analytic Number Theory.* Springer-Verlag, 1976.

[2] P. Flajolet and R. Sedgewick. Digital search trees revisited. *SIAM Journal on Computing*, 15(3):748–767, August 1986.

[3] Ph. Flajolet and G. N. Martin. Probabilistic counting algorithms for data base applications. *Journal of Computer and System Sciences*, 31(2):182–209, October 1985.

[4] Ph. Flajolet and C. Puech. Partial match retrieval of multidimensional data. *Journal of the ACM*, 33(2):371–407, 1986.

[5] P. Kirschenhofer, H. Prodinger, and W. Szpankowski. Multidimensional digital searching and some new parameters in tries. Technical Report CSD-TR-91-052, Purdue University, July 1991.

[6] H. M. Mahmoud. *Evolution of Random Search Trees.* John Wiley, 1992.

[7] G. Tenenbaum. *Introduction à la théorie analytique et probabiliste des nombres.* Institut Élie Cartan, Nancy, 1992.

[8] J. Vitter and Ph. Flajolet. Analysis of Algorithms and Data Structures. In J. van Leeuwen, editor, *Handbook of Theoretical Computer Science*, volume A: Algorithms and Complexity, chapter 9, pages 431–524. North Holland, 1990.

[9] E. T. Whittaker and G. N. Watson. *A Course of Modern Analysis.* Cambridge University Press, fourth edition, 1927. Reprinted 1973.