The Variance for Partial Match Retrievals in k-dimensional Bucket Digital Trees

($\approx$ 1/2 joint with Hsien-Kuei Hwang and Vytas Zacharovas)

Michael Fuchs

Department of Applied Mathematics
National Chiao Tung University

Hsinchu, Taiwan

AofA2010, July 8th, 2010

Bucket Digital Search Tree

Introduced by Coffman and Eve (1970).

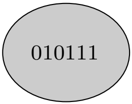
Bucket Digital Search Tree

Introduced by Coffman and Eve (1970).

010111
101011
100001
011011
111110
110111
010011
011110
000100

Bucket Digital Search Tree

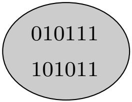
Introduced by Coffman and Eve (1970).



010111
101011
100001
011011
111110
110111
010011
011110
000100

Bucket Digital Search Tree

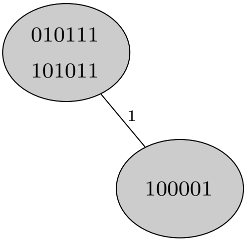
Introduced by Coffman and Eve (1970).



010111
101011
100001
011011
111110
110111
010011
011110
000100

Bucket Digital Search Tree

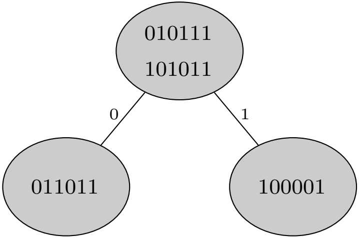
Introduced by Coffman and Eve (1970).



010111
101011
100001
011011
111110
110111
010011
011110
000100

Bucket Digital Search Tree

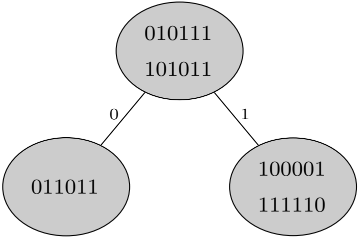
Introduced by Coffman and Eve (1970).



010111
101011
100001
011011
111110
110111
010011
011110
000100

Bucket Digital Search Tree

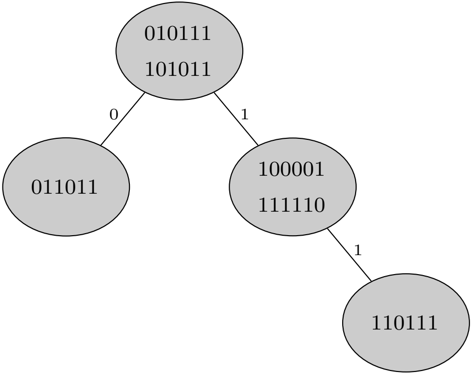
Introduced by Coffman and Eve (1970).



010111
101011
100001
011011
111110
110111
010011
011110
000100

Bucket Digital Search Tree

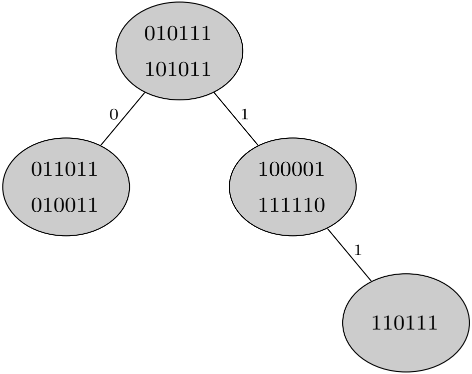
Introduced by Coffman and Eve (1970).



010111
101011
100001
011011
111110
110111
010011
011110
000100

Bucket Digital Search Tree

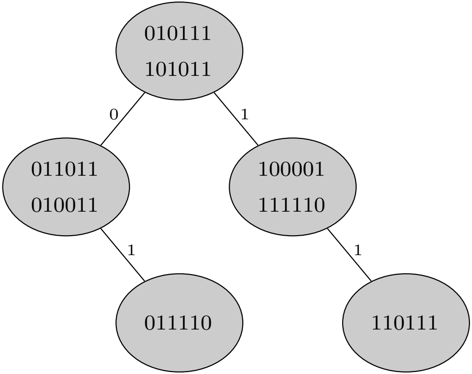
Introduced by Coffman and Eve (1970).



010111
101011
100001
011011
111110
110111
010011
011110
000100

Bucket Digital Search Tree

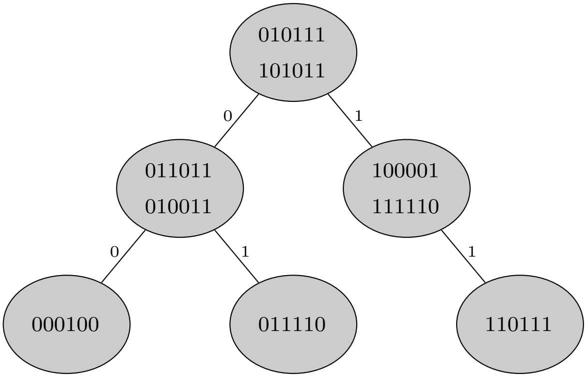
Introduced by Coffman and Eve (1970).



010111
101011
100001
011011
111110
110111
010011
011110
000100

Bucket Digital Search Tree

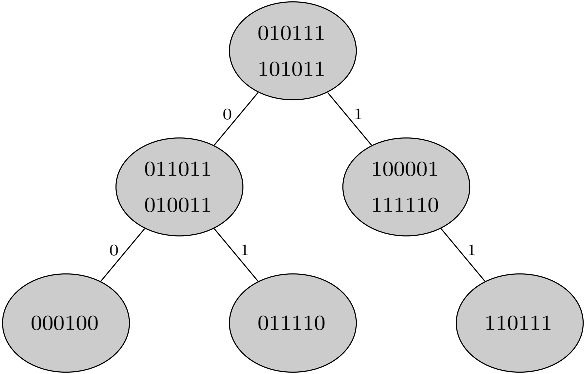
Introduced by Coffman and Eve (1970).



010111
101011
100001
011011
111110
110111
010011
011110
000100

Bucket Digital Search Tree

Introduced by Coffman and Eve (1970).



010111
101011
100001
011011
111110
110111
010011
011110
000100

Random Model: 0-1 are generated independently and equally likely.

Shape Parameters

- **Depth**

  Konheim, Newman, Knuth, Devroye, Louchard, Szpankowski

- **Partial Match Queries**

  Flajolet, Puech, Kirschenhofer, Prodinger, Szpankowski, Schachinger

- **# of Occurrences of Patterns**

  Knuth, Flajolet, Sedgewick, Prodinger, Kirschenhofer

- **Key-Wise Path Length**

  Flajolet, Sedgewick, Prodinger, Kirschenhofer, Szpankowski, Hubalek

- **Node-Wise Path Length**

  Fuchs, Hwang, Zacharovas

Previous Approaches

- **Rice Method**

  Introduced by Flajolet and Sedgewick for digital search trees with bucket size one.

- **Approach of Flajolet and Richmond**

  Introduced for the analysis of bucket digital search trees. Based on Euler transform, Mellin transform, and singularity analysis.

- **Approach via Analytic Depoissonization**

  Introduced by Jacquet & Regnier and Jacquet & Szpankowski. Based on saddle point method and Mellin transform.

- **Schachinger's Approach**

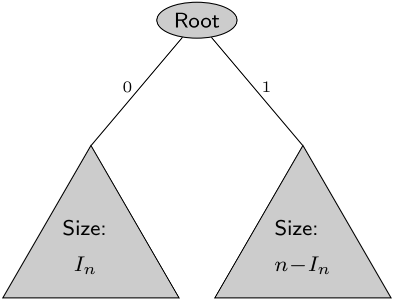
  Largely elementary.

Distributional Recurrence

Shape parameters X_n satisfy the recurrence:

$$X_{n+b} \stackrel{d}{=} X_{I_n} + X^*_{n-I_n} + T_n$$

- $I_n \stackrel{d}{=} \text{Binomial}(n, 1/2)$;
- $X_n \stackrel{d}{=} X^*_n$;
- X_n, X^*_n, I_n independent.
- T_n toll-function.



Poissonization

Moments satisfy the recurrence:

$$f_{n+b} = 2^{1-n} \sum_{j=0}^{n} \binom{n}{j} f_j + g_n.$$

Poissonization

Moments satisfy the recurrence:

$$f_{n+b} = 2^{1-n} \sum_{j=0}^{n} \binom{n}{j} f_j + g_n.$$

Consider Poisson-generating function of f_n and g_n, i.e.,

$$\tilde{f}(z) := e^{-z} \sum_{n \geq 0} f_n \frac{z^n}{n!}, \qquad \tilde{g}(z) := e^{-z} \sum_{n \geq 0} g_n \frac{z^n}{n!}.$$

Poissonization

Moments satisfy the recurrence:

$$f_{n+b} = 2^{1-n} \sum_{j=0}^{n} \binom{n}{j} f_j + g_n.$$

Consider Poisson-generating function of f_n and g_n, i.e.,

$$\tilde{f}(z) := e^{-z} \sum_{n \geq 0} f_n \frac{z^n}{n!}, \qquad \tilde{g}(z) := e^{-z} \sum_{n \geq 0} g_n \frac{z^n}{n!}.$$

Then,

$$\sum_{j=0}^{b} \binom{b}{j} \tilde{f}^{(j)}(z) = 2\tilde{f}(z/2) + \tilde{g}(z).$$

Poissonized Variance

Poisson Heuristic:

$$f_n \text{ sufficiently smooth} \quad \Longrightarrow \quad \tilde{f}(n) \approx f_n.$$

Poissonized Variance

Poisson Heuristic:

$$f_n \text{ sufficiently smooth} \quad \Longrightarrow \quad \tilde{f}(n) \approx f_n.$$

$\tilde{f}_2, \tilde{f}_1$ Poisson-generating functions of second moment and mean.

Poissonized Variance

Poisson Heuristic:

$$f_n \text{ sufficiently smooth} \quad \Longrightarrow \quad \tilde{f}(n) \approx f_n.$$

$\tilde{f}_2, \tilde{f}_1$ Poisson-generating functions of second moment and mean.

Poissonized Variance:

- If mean is sublinear,

$$\tilde{V}(z) = \tilde{f}_2(z) - \tilde{f}_1(z)^2.$$

Poissonized Variance

Poisson Heuristic:

$$f_n \text{ sufficiently smooth} \quad \Longrightarrow \quad \tilde{f}(n) \approx f_n.$$

$\tilde{f}_2, \tilde{f}_1$ Poisson-generating functions of second moment and mean.

Poissonized Variance:

- If mean is sublinear,

$$\tilde{V}(z) = \tilde{f}_2(z) - \tilde{f}_1(z)^2.$$

- If mean is linear,

$$\tilde{V}(z) = \tilde{f}_2(z) - \tilde{f}_1(z)^2 - z\tilde{f}_1'(z)^2.$$

Jacquet-Szpankowski-admissibility (JS-admissibility)

$\tilde{f}(z)$ is called JS-admissible if

$(\mathbf{I})$ Uniformly for $|\arg(z)| \le \epsilon$,

$$\tilde{f}(z) = \mathcal{O}\left(|z|^{\alpha} \log^{\beta} |z|\right),$$

$(\mathbf{O})$ Uniformly for $\epsilon < |\arg(z)| \le \pi$,

$$f(z) := e^{z} \tilde{f}(z) = \mathcal{O}\left(e^{(1-\epsilon)|z|}\right).$$

Jacquet-Szpankowski-admissibility (JS-admissibility)

$\tilde{f}(z)$ is called JS-admissible if

(I) Uniformly for $|\arg(z)| \leq \epsilon$,

$$\tilde{f}(z) = \mathcal{O}\left(|z|^{\alpha} \log^{\beta} |z|\right),$$

(O) Uniformly for $\epsilon < |\arg(z)| \leq \pi$,

$$f(z) := e^{z} \tilde{f}(z) = \mathcal{O}\left(e^{(1-\epsilon)|z|}\right).$$

Theorem (Jacquet and Szpankowski)

If $\tilde{f}(z)$ is JS-admissible, then

$$f_n \sim \tilde{f}(n) - \frac{n}{2}\tilde{f}''(n) + \cdots.$$

Depoissonization

JS-admissibility satisfies closure properties:

(i) $\tilde{f}, \tilde{g}$ JS-admissible, then $\tilde{f} + \tilde{g}$ JS-admissible.

(ii) $\tilde{f}$ JS-admissible, then $\tilde{f}'$ JS-admissible. Etc.

Depoissonization

JS-admissibility satisfies closure properties:

(i) $\tilde{f}, \tilde{g}$ JS-admissible, then $\tilde{f} + \tilde{g}$ JS-admissible.

(ii) $\tilde{f}$ JS-admissible, then $\tilde{f}'$ JS-admissible. Etc.

Proposition

Consider

$$\sum_{j=0}^{b} \binom{b}{j} \tilde{f}^{(j)}(z) = 2\tilde{f}(z/2) + \tilde{g}(z).$$

We have,

$$\tilde{g}(z) \text{ JS-admissible} \quad \Longrightarrow \quad \tilde{f}(z) \text{ JS-admissible.}$$

Laplace-Mellin Approach (i)

We start from,

$$\sum_{j=0}^{b} \binom{b}{j} \tilde{f}^{(j)}(z) = 2\tilde{f}(z/2) + \tilde{g}(z).$$

Laplace-Mellin Approach (i)

We start from,

$$\sum_{j=0}^{b} \binom{b}{j} \tilde{f}^{(j)}(z) = 2\tilde{f}(z/2) + \tilde{g}(z).$$

Applying Laplace transform,

$$(s+1)^b \mathscr{L}[\tilde{f}(z); s] = 4\mathscr{L}[\tilde{f}(z); 2s] + \mathscr{L}[\tilde{g}(z); s] + p(s).$$

with $p(s)$ a polynomial.

Laplace-Mellin Approach (i)

We start from,

$$\sum_{j=0}^{b} \binom{b}{j} \tilde{f}^{(j)}(z) = 2\tilde{f}(z/2) + \tilde{g}(z).$$

Applying Laplace transform,

$$(s+1)^b \mathscr{L}[\tilde{f}(z); s] = 4\mathscr{L}[\tilde{f}(z); 2s] + \mathscr{L}[\tilde{g}(z); s] + p(s).$$

with $p(s)$ a polynomial.

Define,

$$Q(s) := \sum_{l \geq 1} \left(1 - \frac{s}{2^l}\right)$$

and $Q_\infty := Q(1)$.

Laplace-Mellin Approach (ii)

Set

$$\bar{f}(s) := \frac{\mathscr{L}[\tilde{f}(z); s]}{Q(-s)^b}, \qquad \bar{g}(s) := \frac{\mathscr{L}[\tilde{g}(z); s] + p(s)}{Q(-2s)^b}.$$

Laplace-Mellin Approach (ii)

Set

$$\bar{f}(s) := \frac{\mathscr{L}[\tilde{f}(z); s]}{Q(-s)^b}, \qquad \bar{g}(s) := \frac{\mathscr{L}[\tilde{g}(z); s] + p(s)}{Q(-2s)^b}.$$

Then,

$$\bar{f}(s) = 4\bar{f}(2s) + \bar{g}(s).$$

Laplace-Mellin Approach (ii)

Set

$$\bar{f}(s) := \frac{\mathscr{L}[\tilde{f}(z); s]}{Q(-s)^b}, \qquad \bar{g}(s) := \frac{\mathscr{L}[\tilde{g}(z); s] + p(s)}{Q(-2s)^b}.$$

Then,

$$\bar{f}(s) = 4\bar{f}(2s) + \bar{g}(s).$$

Applying Mellin transform,

$$\mathscr{M}[\bar{f}(s); \omega] = \frac{\mathscr{M}[\bar{g}(s); \omega]}{1 - 2^{2-\omega}}.$$

Laplace-Mellin Approach (ii)

Set

$$\bar{f}(s) := \frac{\mathscr{L}[\tilde{f}(z); s]}{Q(-s)^b}, \qquad \bar{g}(s) := \frac{\mathscr{L}[\tilde{g}(z); s] + p(s)}{Q(-2s)^b}.$$

Then,

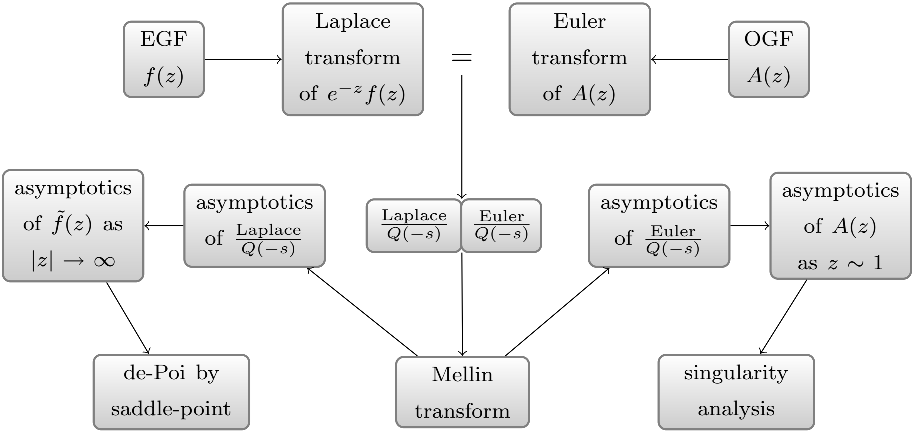
$$\bar{f}(s) = 4\bar{f}(2s) + \bar{g}(s).$$

Applying Mellin transform,

$$\mathscr{M}[\bar{f}(s); \omega] = \frac{\mathscr{M}[\bar{g}(s); \omega]}{1 - 2^{2-\omega}}.$$

From this, an asymptotic expansion of $\tilde{f}(z)$ as $z \to \infty$ is obtained via inverse Mellin transform and inverse Laplace transform.

Our Approach vs. Flajolet-Richmond



Variance for Key-Wise Path Length

Theorem

We have,

$$\mathrm{Var}(X_n) \sim nP_2(\log_2 n),$$

where $P_2(z)$ is a one-periodic function with Fourier coefficients

$$\frac{1}{L\Gamma(2+2\pi ir/L)} \int_0^\infty \frac{s^{1+2\pi ir/L}}{Q(-2s)^b} \int_0^\infty e^{-zs}\tilde{h}(z)\mathrm{d}z\mathrm{d}s$$

with $L := \log 2$.

Here, $\tilde{h}(z)$ is a function of the Poisson generating function of the mean.

Fourier Coefficients

Theorem

For $b = 1$,

$$\frac{Q_\infty}{L\Gamma(2+2\pi ir/L)} \sum_{j_1,j_2,j_3\geq 0} \frac{(-1)^{j_1}2^{-\binom{j_1+1}{2}+2\pi irj_1/L}}{Q_{j_1}Q_{j_2}Q_{j_3}2^{j_2+j_3}}\varphi(2+2\pi ir/L; 2^{-j_1-j_2}+2^{-j_1-j_3})$$

with $Q_j = \prod_{l=1}^{j}(1-2^l)$ *and*

$$\varphi(\omega; x) = \frac{\pi(1+x^{\omega-2}((\omega-2)x+1-\omega))}{(x-1)^2\sin(\pi\omega)}.$$

$$
\begin{aligned}
&-\frac{28}{3L}-\frac{39}{4}-2\sum_{l\ge1}\frac{l2^l}{(2^l-1)^2}+\frac{2}{L}\sum_{l\ge1}\frac{1}{2^l-1}+\frac{\pi^2}{2L^2}+\frac{2}{L^2}\\
&-\frac{2}{L}\sum_{l\ge3}\frac{(-1)^{l+1}(l-5)}{(l+1)l(l-1)(2^l-1)}\\
&+\frac{2}{L}\sum_{l\ge1}(-1)^l2^{-\binom{l+1}{2}}\left(\frac{L(1-2^{-l+1})/2-1}{1-2^{-l}}-\sum_{r\ge2}\frac{(-1)^{r+1}}{r(r-1)(2^{r+l}-1)}\right)\\
&-\frac{2Q(1)}{L}+\sum_{l\ge2}\frac{1}{2^lQ_l}\sum_{r\ge0}\frac{(-1)^r2^{-\binom{r+1}{2}}}{Q_r}Q_{r+l-2}\cdot\\
&\cdot\left(-\sum_{j\ge1}\frac{1}{2^{j+r+l+2}-1}\left(2^{l+1}-2l-4+2\sum_{i=2}^{l-1}\binom{l+1}{i}\frac{1}{2^{r+i-1}-1}\right)\right.\\
&\quad+\frac{2}{(1-2^{-l-r})^2}+\frac{2l+2}{(1-2^{1-l-r})^2}-\frac{2}{L}\frac{1}{1-2^{1-l-r}}+\frac{2}{L}\sum_{j=1}^{l+1}\binom{l+1}{j}\frac{1}{2^{r+j}-1}\\
&\quad\left.-2\sum_{j=2}^{l+1}\binom{l+1}{j}\frac{1}{2^{r+j-1}-1}+\frac{2}{L}\sum_{j=0}^{l+1}\binom{l+1}{j}\sum_{i\ge1}\frac{(-1)^i}{(i+1)(2^{r+j+i}-1)}\right)\\
&+\sum_{l\ge3}\sum_{r=2}^{l-1}\binom{l+1}{r}\frac{Q_{r-2}Q_{l-r-1}}{2^lQ_l}\sum_{j\ge l+1}\frac{1}{2^j-1}-2[FH]_0-[F^2]_0.
\end{aligned}
$$

k-d Bucket Digital Search Trees

Let $R_1, \ldots, R_n$ be k-dimensional data, i.e., R_i consist of k 0-1 strings

$$\begin{aligned} R_{i,1} &= \left(R_{i,1}^{[1]}, R_{i,1}^{[2]}, R_{i,1}^{[3]}, \ldots\right), \\ &\vdots \\ R_{i,k} &= \left(R_{i,k}^{[1]}, R_{i,k}^{[2]}, R_{i,k}^{[3]}, \ldots\right). \end{aligned}$$

k-d Bucket Digital Search Trees

Let $R_1, \ldots, R_n$ be k-dimensional data, i.e., R_i consist of k 0-1 strings

$$R_{i,1} = \left(R_{i,1}^{[1]}, R_{i,1}^{[2]}, R_{i,1}^{[3]}, \ldots\right),$$
$$\vdots$$
$$R_{i,k} = \left(R_{i,k}^{[1]}, R_{i,k}^{[2]}, R_{i,k}^{[3]}, \ldots\right).$$

Shuffling yields

$$\tilde{R}_i = \left(R_{i,1}^{[1]}, \ldots, R_{i,k}^{[1]}, R_{i,1}^{[2]}, \ldots, R_{i,k}^{[2]}, \ldots\right).$$

Use $\tilde{R}_1, \ldots, \tilde{R}_n$ to construct the usual bucket digital search tree.

k-d Bucket Digital Search Trees

Let $R_1, \ldots, R_n$ be k-dimensional data, i.e., R_i consist of k 0-1 strings

$$R_{i,1} = \left(R_{i,1}^{[1]}, R_{i,1}^{[2]}, R_{i,1}^{[3]}, \ldots\right),$$
$$\vdots$$
$$R_{i,k} = \left(R_{i,k}^{[1]}, R_{i,k}^{[2]}, R_{i,k}^{[3]}, \ldots\right).$$

Shuffling yields

$$\tilde{R}_i = \left(R_{i,1}^{[1]}, \ldots, R_{i,k}^{[1]}, R_{i,1}^{[2]}, \ldots, R_{i,k}^{[2]}, \ldots\right).$$

Use $\tilde{R}_1, \ldots, \tilde{R}_n$ to construct the usual bucket digital search tree.

$\longrightarrow$ k-d bucket digital search tree.

Partial Match Queries

- Let $Q = (Q_1, \ldots, Q_k)$ be a *partial match query*, where Q_i is either a 0-1 string or a an undefined string.

Partial Match Queries

- Let $Q = (Q_1, \ldots, Q_k)$ be a *partial match query*, where Q_i is either a 0-1 string or a an undefined string.

- Construct a shuffled record $\tilde{Q}$ as before.

Partial Match Queries

- Let $Q = (Q_1, \ldots, Q_k)$ be a *partial match query*, where Q_i is either a 0-1 string or a an undefined string.
- Construct a shuffled record $\tilde{Q}$ as before.
- Cost:=# nodes visited when $\tilde{Q}$ is used as search query.

Partial Match Queries

- Let $Q = (Q_1, \dots, Q_k)$ be a *partial match query*, where Q_i is either a 0-1 string or a an undefined string.
- Construct a shuffled record $\tilde{Q}$ as before.
- Cost:$=\#$ nodes visited when $\tilde{Q}$ is used as search query.
- The cost only depends on the *partial match pattern* $q \in \{S, \star\}^k$, where

$$q_i = \begin{cases} S, & \text{if } Q_i \text{ is a } 0\text{-}1 \text{ string;} \\ \star, & \text{if } Q_i \text{ is unspecified.} \end{cases}$$

Partial Match Queries

- Let $Q = (Q_1, \ldots, Q_k)$ be a *partial match query*, where Q_i is either a 0-1 string or a an undefined string.
- Construct a shuffled record $\tilde{Q}$ as before.
- Cost:$=\#$ nodes visited when $\tilde{Q}$ is used as search query.
- The cost only depends on the *partial match pattern* $q \in \{S, \star\}^k$, where

$$q_i = \begin{cases} S, & \text{if } Q_i \text{ is a } 0\text{-}1 \text{ string;} \\ \star, & \text{if } Q_i \text{ is unspecified.} \end{cases}$$

- Cost will be denoted by $X_{q,n}$, where n is the size of the tree.

Distributional Recurrence

Similar as before,

$$X_{q,n+b} \stackrel{d}{=} \begin{cases} X_{q',I_n} + X^*_{q',n-I_n} + 1, & \text{if } q = (\star, \dots); \\ X_{q',I_n} + 1, & \text{if } q = (S, \dots), \end{cases}$$

where q' denotes the cyclic shift to the left.

Distributional Recurrence

Similar as before,

$$X_{q,n+b} \stackrel{d}{=} \begin{cases} X_{q',I_n} + X^*_{q',n-I_n} + 1, & \text{if } q = (\star, \ldots); \\ X_{q',I_n} + 1, & \text{if } q = (S, \ldots), \end{cases}$$

where q' denotes the cyclic shift to the left.

Applying Poisson generating function gives for the moments

$$\sum_{j=0}^{b} \binom{b}{j} \tilde{f}^{(j)}_{q^{(l)}}(z) = \delta_{l+1} \tilde{f}_{q^{(l+1)}}(z/2) + \tilde{g}_{q^{(l)}}(z),$$

where $q^{(l)}$ is the cyclic shift applied l times and $\delta_{l+1} \in \{1, 2\}$.

Mean Value

Theorem (Kirschenhofer and Prodinger 1994)

For $b = 1$ and u the number of unspecified coordinates,

$$\mathbf{E}(X_{q,n}) \sim u^{u/k} P_1(\log_2 n^{1/k}),$$

where $P_1(z)$ is one-periodic with Fourier coefficients

$$\frac{\mathscr{M}[1/(sQ(-2s)^b); \omega_r]}{kL\Gamma(1+\omega_r)} \sum_{l=0}^{k-1} \delta_1 \dots \delta_l 2^{-\omega_r l}$$

with $w_r = u/k + 2\pi i r/(kL)$.

Depoissonization

Since mean value is sublinear, Poissonized variance is given by

$$\tilde{V}_q(z) := \tilde{f}_{q,2}(z) - \tilde{f}_{q,1}(z)^2.$$

Depoissonization

Since mean value is sublinear, Poissonized variance is given by

$$\tilde{V}_q(z) := \tilde{f}_{q,2}(z) - \tilde{f}_{q,1}(z)^2.$$

Depoissonization is done via the following result + closure properties.

Proposition

Consider

$$\sum_{j=0}^{b} \binom{b}{j} \tilde{f}^{(j)}_{q^{(l)}}(z) = \delta_{l+1} \tilde{f}_{q^{(l+1)}}(z/2) + \tilde{g}_{q^{(l)}}(z).$$

We have,

$$\tilde{g}_{q^{(l)}}(z) \text{ JS-admissible} \quad \Longrightarrow \quad \tilde{f}_{q^{(l)}}(z) \text{ JS-admissible.}$$

Mean, Second Moment and Variance

We have,

$$\sum_{j=0}^{b} \binom{b}{j} \tilde{f}_{q,1}^{(j)}(z) = \delta_1 \tilde{f}_{q',1}(z/2) + 1$$

Mean, Second Moment and Variance

We have,

$$\sum_{j=0}^{b}\binom{b}{j}\tilde{f}_{q,1}^{(j)}(z)=\delta_1\tilde{f}_{q',1}(z/2)+1$$

and

$$\sum_{j=0}^{b}\binom{b}{j}\tilde{f}_{q,2}^{(j)}(z)=\delta_1\tilde{f}_{q',2}(z/2)+\tilde{g}_{q,2}(z).$$

where

$$\tilde{g}_{q,2}(z)=\begin{cases}4\tilde{f}_{q',1}(z/2)+2\tilde{f}_{q',1}(z/2)^2+1, & \text{if } q^{(l)}=(\star,\ldots);\\ 2\tilde{f}_{q',1}(z/2)+1, & \text{if } q^{(l)}=(S,\ldots).\end{cases}$$

Mean, Second Moment and Variance

We have,

$$\sum_{j=0}^{b}\binom{b}{j}\tilde{f}_{q,1}^{(j)}(z)=\delta_1\tilde{f}_{q',1}(z/2)+1$$

and

$$\sum_{j=0}^{b}\binom{b}{j}\tilde{f}_{q,2}^{(j)}(z)=\delta_1\tilde{f}_{q',2}(z/2)+\tilde{g}_{q,2}(z).$$

where

$$\tilde{g}_{q,2}(z)=\begin{cases}4\tilde{f}_{q',1}(z/2)+2\tilde{f}_{q',1}(z/2)^2+1, & \text{if } q^{(l)}=(\star,\dots);\\ 2\tilde{f}_{q',1}(z/2)+1, & \text{if } q^{(l)}=(S,\dots).\end{cases}$$

$\tilde{f}_{q,1}(z), \tilde{f}_{q,2}(z)$ are JS-admissible and hence $\tilde{V}_q(n)\sim \mathrm{Var}(X_n)$.

Laplace-Mellin Approach (i)

We start from,

$$\sum_{j=0}^{b} \binom{b}{j} \tilde{f}_{q^{(l)}}^{(j)}(z) = \delta_{l+1} \tilde{f}_{q^{(l+1)}}(z/2) + \tilde{g}_{q^{(l)}}(z).$$

Laplace-Mellin Approach (i)

We start from,

$$\sum_{j=0}^{b} \binom{b}{j} \tilde{f}^{(j)}_{q^{(l)}}(z) = \delta_{l+1} \tilde{f}_{q^{(l+1)}}(z/2) + \tilde{g}_{q^{(l)}}(z).$$

Applying Laplace transform,

$$(s+1)^b [\tilde{f}_{q^{(l)}}(z); s] = 2\delta_{l+1} \mathscr{L}[\tilde{f}_{q^{(l+1)}}(z); 2s] + \mathscr{L}[\tilde{g}_{q^{(l)}}(z); s] + p(s),$$

where $p(s)$ is some polynomial.

Laplace-Mellin Approach (i)

We start from,

$$\sum_{j=0}^{b} \binom{b}{j} \tilde{f}_{q^{(l)}}^{(j)}(z) = \delta_{l+1} \tilde{f}_{q^{(l+1)}}(z/2) + \tilde{g}_{q^{(l)}}(z).$$

Applying Laplace transform,

$$(s+1)^b [\tilde{f}_{q^{(l)}}(z); s] = 2\delta_{l+1} \mathscr{L}[\tilde{f}_{q^{(l+1)}}(z); 2s] + \mathscr{L}[\tilde{g}_{q^{(l)}}(z); s] + p(s),$$

where $p(s)$ is some polynomial.

Next normalize,

$$\bar{f}_{q^{(l)}}(s) = \frac{\mathscr{M}[\tilde{f}_{q^{(l)}}(z); s]}{Q(-s)^b}, \qquad \bar{g}_{q^{(l)}}(s) = \frac{\mathscr{M}[\tilde{g}_{q^{(l)}}(z); s] + p(s)}{Q(-2s)^b}.$$

Laplace-Mellin Approach (ii)

This yields,

$$\bar{f}_{q^{(l)}}(s) = 2\delta_{l+1}\bar{f}_{q^{(l+1)}}(2s) + \bar{g}_{q^{(l)}}(s).$$

Laplace-Mellin Approach (ii)

This yields,

$$\bar{f}_{q^{(l)}}(s) = 2\delta_{l+1}\bar{f}_{q^{(l+1)}}(2s) + \bar{g}_{q^{(l)}}(s).$$

By iteration

$$\bar{f}_q(s) = 2^{k+u}\bar{f}_q(2^k s) + \sum_{l=0}^{k-1} \delta_1 \cdots \delta_l 2^l \bar{g}_{q^{(l)}}(2^l s).$$

Laplace-Mellin Approach (ii)

This yields,

$$\bar{f}_{q^{(l)}}(s) = 2\delta_{l+1}\bar{f}_{q^{(l+1)}}(2s) + \bar{g}_{q^{(l)}}(s).$$

By iteration

$$\bar{f}_q(s) = 2^{k+u}\bar{f}_q(2^k s) + \sum_{l=0}^{k-1} \delta_1 \cdots \delta_l 2^l \bar{g}_{q^{(l)}}(2^l s).$$

Applying Mellin transform

$$\mathscr{M}[\bar{f}_q(s);\omega] = \frac{1}{1-2^{k-\omega k+u}} \sum_{l=0}^{k-1} \delta_1 \cdots \delta_l 2^{l-\omega l} \mathscr{M}[\bar{g}_{q^{(l)}}(s);\omega].$$

Laplace-Mellin Approach (ii)

This yields,

$$\bar{f}_{q^{(l)}}(s) = 2\delta_{l+1}\bar{f}_{q^{(l+1)}}(2s) + \bar{g}_{q^{(l)}}(s).$$

By iteration

$$\bar{f}_q(s) = 2^{k+u}\bar{f}_q(2^k s) + \sum_{l=0}^{k-1} \delta_1 \cdots \delta_l 2^l \bar{g}_{q^{(l)}}(2^l s).$$

Applying Mellin transform

$$\mathscr{M}[\bar{f}_q(s);\omega] = \frac{1}{1-2^{k-\omega k+u}} \sum_{l=0}^{k-1} \delta_1 \cdots \delta_l 2^{l-\omega l} \mathscr{M}[\bar{g}_{q^{(l)}}(s);\omega].$$

The analysis is completed by Inverse Mellin + Laplace.

Variance

Theorem

For u *the number of unspecified coordinates,*

$$\mathrm{Var}(X_{q,n}) \sim n^{u/k} P_2(\log_2 n^{1/k}),$$

where $P_2(z)$ *is one-periodic with Fourier coefficients*

$$\frac{1}{kL\Gamma(1+\omega_r)} \sum_{l=0}^{k-1} \delta_1 \dots \delta_l 2^{-\omega_r l} \int_0^\infty \frac{s^{\omega_r}}{Q(-2s)^b} \left(\int_0^\infty e^{-zs} \tilde{h}_q^l(z) \mathrm{d}z + p(s) \right) \mathrm{d}s$$

with $p(s)$ *a polynomial and* $\tilde{h}_q^l(z)$ *a function of the Poisson generating function of the mean.*

Fourier coefficients

Corollary

For $b = 1$,

$$\frac{1}{kLQ_\infty\Gamma(1+\omega_r)}\sum_{l=0}^{k-1}\delta_1\cdots\delta_l 2^{-\omega_r l}$$

$$\sum_{j_1,j_2,j_3\geq 0}\frac{(-1)^{j_1}\bar{\delta}_{q^{(l)},j_2}\bar{\delta}_{q^{(l)},j_3}2^{-\binom{j_1}{2}+(1-\omega_r)j_1}}{2^{j_2+j_3}Q_{j_1}Q_{j_2}Q_{j_3}}\varphi(\omega_r;2^{j_1-j_2}+2^{j_1-j_3})$$

with

$$\bar{\delta}_{q,j}=\sum_{l\geq 0}\frac{(-1)^l 2^{-\binom{l+1}{2}}}{Q_l}\prod_{h=1}^{l+j}\delta_h.$$

and $\varphi(\omega;x)=\pi(x^\omega-1)/(\sin(\pi\omega)(x-1))$.

Variants of Digital Search Trees

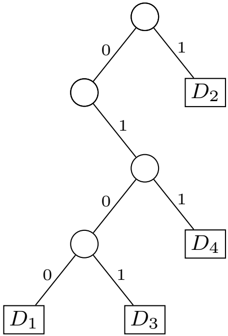
- *Tries*: data is only stored in the leaves.

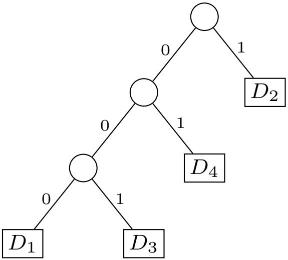
Variants of Digital Search Trees

- *Tries*: data is only stored in the leaves.
- *PATRICIA tries*: one-way branching is suppressed.

Variants of Digital Search Trees

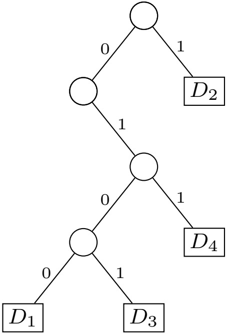
- *Tries*: data is only stored in the leaves.
- *PATRICIA tries*: one-way branching is suppressed.

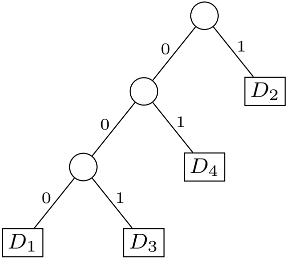




Variants of Digital Search Trees

- *Tries*: data is only stored in the leaves.
- *PATRICIA tries*: one-way branching is suppressed.





$X_{q,n} = \#$ of internal nodes visited by the query.

Previous Results for Tries

- Flajolet and Puech showed that for $b = 1$

$$\mathbf{E}(X_{q,n}) \sim n^{u/k} P_1(\log_2 n^{1/k}).$$

Previous Results for Tries

- Flajolet and Puech showed that for $b = 1$

$$\mathbf{E}(X_{q,n}) \sim n^{u/k} P_1(\log_2 n^{1/k}).$$

- Kirschenhofer and Prodinger generalized this to arbitrary b.

Previous Results for Tries

- Flajolet and Puech showed that for $b = 1$

$$\mathbf{E}(X_{q,n}) \sim n^{u/k} P_1(\log_2 n^{1/k}).$$

- Kirschenhofer and Prodinger generalized this to arbitrary b.

- Kirschenhofer, Prodinger and Szpankowski showed that for $k = 2$ and $b = 1$

$$\mathrm{Var}(X_{q,n}) \sim n^{u/k} P_2(\log_2 n^{1/k}).$$

They conjectured that the same result also holds for general k.

Previous Results for Tries

- Flajolet and Puech showed that for $b = 1$

$$\mathbf{E}(X_{q,n}) \sim n^{u/k} P_1(\log_2 n^{1/k}).$$

- Kirschenhofer and Prodinger generalized this to arbitrary b.

- Kirschenhofer, Prodinger and Szpankowski showed that for $k = 2$ and $b = 1$

$$\mathrm{Var}(X_{q,n}) \sim n^{u/k} P_2(\log_2 n^{1/k}).$$

  They conjectured that the same result also holds for general k.

- Schachinger proved the above conjecture.

Variance for Tries

Theorem

For u the number of unspecified coordinates,

$$\mathrm{Var}(X_{q,n}) \sim n^{u/k} P_2(\log_2 n^{1/k}),$$

where $P_2(z)$ is one-periodic with Fourier coefficients

$$\frac{\Gamma(-\omega_r)}{kL}\left(\delta(2^{-\omega_r})\left(\binom{-\omega_r+b}{b} - 2^{\omega_r}\sum_{j_1,j_2=0}^{b}\binom{j_1+j_2}{j_1}\binom{-\omega_r+j_1+j_2}{j_1+j_2}2^{-j_1-j_2}\right)\right.$$
$$\left. - \sum_{l\geq b+1}\binom{-l+b}{b}\binom{-\omega_r+l+b}{b}\binom{\omega_r}{l}\frac{2^{1-l}\sigma(2^{-\omega_r},2^{-l}))}{1-2^{-lk+u}}\right),$$

where

$$\delta(z) = \sum_{j=0}^{k-1}\delta_1\cdots\delta_j z^j, \qquad \sigma(z_1,z_2) = \sum_{j_1,j_2=0}^{k-1}\delta_1\cdots\delta_{j_1+j_2+1} z_1^{j_1} z_2^{j_2}.$$