MADALGO
CENTER FOR MASSIVE DATA ALGORITHMICS

Hardness of Several String Indexing Problems

Jesper Sindahl Nielsen

Joint work with:
Kasper Green Larsen, J. Ian Munro and Sharma V. Thankachan

Danmarks Grundforskningsfond
Danish National Research Foundation

AARHUS UNIVERSITY

Problems

Problems

1 Preprocess documents $\mathcal{D} = \{d_1, d_2, \ldots, d_D\}$ with $\sum_{i=1}^{D} |d_i| = n$, to answer queries: given two patterns P_1 and P_2 report all documents where both P_1 and P_2 occur in d_i. (two patterns).

Problems

1. Preprocess documents $\mathcal{D} = \{d_1, d_2, \ldots, d_D\}$ with $\sum_{i=1}^{D} |d_i| = n$, to answer queries: given two patterns P_1 and P_2 report all documents where both P_1 and P_2 occur in d_i. (two patterns).
2. Preprocess documents $\mathcal{D} = \{d_1, d_2, \ldots, d_D\}$ with $\sum_{i=1}^{D} |d_i| = n$, to answer queries: given two patterns P^+ and P^- report all documents where P^+ occurs and P^- does not occur (excluded pattern).

Problems

1. Preprocess documents $\mathcal{D} = \{d_1, d_2, \ldots, d_D\}$ with $\sum_{i=1}^{D} |d_i| = n$, to answer queries: given two patterns P_1 and P_2 report all documents where both P_1 and P_2 occur in d_i. (two patterns).
2. Preprocess documents $\mathcal{D} = \{d_1, d_2, \ldots, d_D\}$ with $\sum_{i=1}^{D} |d_i| = n$, to answer queries: given two patterns P^+ and P^- report all documents where P^+ occurs and P^- does not occur (excluded pattern).

Previous work

Authors	Query time	Space

Problems

1. Preprocess documents $\mathcal{D} = \{d_1, d_2, \ldots, d_D\}$ with $\sum_{i=1}^{D} |d_i| = n$, to answer queries: given two patterns P_1 and P_2 report all documents where both P_1 and P_2 occur in d_i. (two patterns).
2. Preprocess documents $\mathcal{D} = \{d_1, d_2, \ldots, d_D\}$ with $\sum_{i=1}^{D} |d_i| = n$, to answer queries: given two patterns P^+ and P^- report all documents where P^+ occurs and P^- does not occur (excluded pattern).

Previous work

Authors	Query time	Space
Fischer et al. 2012	$O(\sqrt{n} + k)$	$O(n^{3/2})$ bits

Problems

1 Preprocess documents $\mathcal{D} = \{d_1, d_2, \ldots, d_D\}$ with $\sum_{i=1}^{D} |d_i| = n$, to answer queries: given two patterns P_1 and P_2 report all documents where both P_1 and P_2 occur in d_i. (two patterns).

2 Preprocess documents $\mathcal{D} = \{d_1, d_2, \ldots, d_D\}$ with $\sum_{i=1}^{D} |d_i| = n$, to answer queries: given two patterns P^+ and P^- report all documents where P^+ occurs and P^- does not occur (excluded pattern).

Previous work

Authors	Query time	Space
Fischer et al. 2012	$O(\sqrt{n} + k)$	$O(n^{3/2})$ bits
Hon et al. 2012	$O(\sqrt{nk \log n} \log^2 n)$	$O(n)$ words

Problems

1. Preprocess documents $\mathcal{D} = \{d_1, d_2, \ldots, d_D\}$ with $\sum_{i=1}^{D} |d_i| = n$, to answer queries: given two patterns P_1 and P_2 report all documents where both P_1 and P_2 occur in d_i. (two patterns).
2. Preprocess documents $\mathcal{D} = \{d_1, d_2, \ldots, d_D\}$ with $\sum_{i=1}^{D} |d_i| = n$, to answer queries: given two patterns P^+ and P^- report all documents where P^+ occurs and P^- does not occur (excluded pattern).

Previous work

Authors	Query time	Space
Fischer et al. 2012	$O(\sqrt{n} + k)$	$O(n^{3/2})$ bits
Hon et al. 2012	$O(\sqrt{nk \log n} \log^2 n)$	$O(n)$ words
Hon et al. 2012	$O(\sqrt{n})$ (counting)	$O(n)$ words

Problems

1. Preprocess documents $\mathcal{D} = \{d_1, d_2, \ldots, d_D\}$ with $\sum_{i=1}^{D} |d_i| = n$, to answer queries: given two patterns P_1 and P_2 report all documents where both P_1 and P_2 occur in d_i. (two patterns).
2. Preprocess documents $\mathcal{D} = \{d_1, d_2, \ldots, d_D\}$ with $\sum_{i=1}^{D} |d_i| = n$, to answer queries: given two patterns P^+ and P^- report all documents where P^+ occurs and P^- does not occur (excluded pattern).

Previous work

Authors	Query time	Space
Fischer et al. 2012	$O(\sqrt{n} + k)$	$O(n^{3/2})$ bits
Hon et al. 2012	$O(\sqrt{nk \log n} \log^2 n)$	$O(n)$ words
Hon et al. 2012	$O(\sqrt{n})$ (counting)	$O(n)$ words
Our result	$O(\sqrt{nk} \log^{1/2+\varepsilon} n)$	$O(n)$ words

Problems

1. Preprocess documents $\mathcal{D} = \{d_1, d_2, \ldots, d_D\}$ with $\sum_{i=1}^{D} |d_i| = n$, to answer queries: given two patterns P_1 and P_2 report all documents where both P_1 and P_2 occur in d_i. (two patterns).
2. Preprocess documents $\mathcal{D} = \{d_1, d_2, \ldots, d_D\}$ with $\sum_{i=1}^{D} |d_i| = n$, to answer queries: given two patterns P^+ and P^- report all documents where P^+ occurs and P^- does not occur (excluded pattern).

Previous work

Authors	Query time	Space
Fischer et al. 2012	$O(\sqrt{n} + k)$	$O(n^{3/2})$ bits
Hon et al. 2012	$O(\sqrt{nk \log n} \log^2 n)$	$O(n)$ words
Hon et al. 2012	$O(\sqrt{n})$ (counting)	$O(n)$ words
Our result	$O(\sqrt{nk} \log^{1/2+\varepsilon} n)$	$O(n)$ words

We prove it is at least as hard as boolean matrix multiplication

Outline

- Boolean Matrix Multiplication
- Reduction to Two-Patterns Problem
- Summary

Boolean Matrix Multiplication

Boolean Matrix Multiplication

Integer matrix multiplication:

Boolean Matrix Multiplication

Integer matrix multiplication:

$$\begin{bmatrix} 5 & 1 & 7 \\ -10 & 5 & 5 \\ 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} 1 & 4 & 0 \\ -3 & 2 & -8 \\ 0 & 4 & 3 \end{bmatrix} = \begin{bmatrix} & & \\ & & \\ & & \end{bmatrix}$$

Boolean Matrix Multiplication

Integer matrix multiplication:

$$\begin{bmatrix} 5 & 1 & 7 \\ -10 & 5 & 5 \\ 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} 1 & 4 & 0 \\ -3 & 2 & -8 \\ 0 & 4 & 3 \end{bmatrix} = \begin{bmatrix} \square & & \\ & & \\ & & \end{bmatrix}$$

Boolean Matrix Multiplication

Integer matrix multiplication:

$$\begin{bmatrix} 5 & 1 & 7 \\ -10 & 5 & 5 \\ 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} 1 & 4 & 0 \\ -3 & 2 & -8 \\ 0 & 4 & 3 \end{bmatrix} = \begin{bmatrix} \square & & \\ & & \\ & & \end{bmatrix}$$

$5 \times 1 + 1 \times (-3) + 7 \times 0 = 2$

Boolean Matrix Multiplication

Integer matrix multiplication:

$$\begin{bmatrix} 5 & 1 & 7 \\ -10 & 5 & 5 \\ 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} 1 & 4 & 0 \\ -3 & 2 & -8 \\ 0 & 4 & 3 \end{bmatrix} = \begin{bmatrix} 2 & & \\ & & \\ & & \end{bmatrix}$$

$5 \times 1 + 1 \times (-3) + 7 \times 0 = 2$

Boolean Matrix Multiplication

Integer matrix multiplication:

$$\begin{bmatrix} \boxed{5 \quad 1 \quad 7} \\ -10 \quad 5 \quad 5 \\ 0 \quad 0 \quad 1 \end{bmatrix} \times \begin{bmatrix} 1 & \boxed{4} & 0 \\ -3 & \boxed{2} & -8 \\ 0 & \boxed{4} & 3 \end{bmatrix} = \begin{bmatrix} 2 & \boxed{} & \\ & & \\ & & \end{bmatrix}$$

Boolean Matrix Multiplication

Integer matrix multiplication:

$$\begin{bmatrix} 5 & 1 & 7 \\ -10 & 5 & 5 \\ 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} 1 & 4 & 0 \\ -3 & 2 & -8 \\ 0 & 4 & 3 \end{bmatrix} = \begin{bmatrix} 2 & \square & \\ & & \\ & & \end{bmatrix}$$

$5 \times 4 + 1 \times 2 + 7 \times 4 = 50$

Boolean Matrix Multiplication

Integer matrix multiplication:

$$\begin{bmatrix} 5 & 1 & 7 \\ -10 & 5 & 5 \\ 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} 1 & 4 & 0 \\ -3 & 2 & -8 \\ 0 & 4 & 3 \end{bmatrix} = \begin{bmatrix} 2 & 50 & \\ & & \\ & & \end{bmatrix}$$

$5 \times 4 + 1 \times 2 + 7 \times 4 = 50$

Boolean Matrix Multiplication

Integer matrix multiplication:

$$\begin{bmatrix} 5 & 1 & 7 \\ -10 & 5 & 5 \\ 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} 1 & 4 & 0 \\ -3 & 2 & -8 \\ 0 & 4 & 3 \end{bmatrix} = \begin{bmatrix} 2 & 50 & 13 \\ -25 & -10 & -25 \\ 0 & 4 & 3 \end{bmatrix}$$

$5 \times 4 + 1 \times 2 + 7 \times 4 = 50$

Boolean Matrix Multiplication

Integer matrix multiplication:

$$\begin{bmatrix} 5 & 1 & 7 \\ -10 & 5 & 5 \\ 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} 1 & 4 & 0 \\ -3 & 2 & -8 \\ 0 & 4 & 3 \end{bmatrix} = \begin{bmatrix} 2 & 50 & 13 \\ -25 & -10 & -25 \\ 0 & 4 & 3 \end{bmatrix}$$

$5 \times 4 + 1 \times 2 + 7 \times 4 = 50$

Boolean matrix multiplication:

Boolean Matrix Multiplication

Integer matrix multiplication:

$$\begin{bmatrix} 5 & 1 & 7 \\ -10 & 5 & 5 \\ 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} 1 & 4 & 0 \\ -3 & 2 & -8 \\ 0 & 4 & 3 \end{bmatrix} = \begin{bmatrix} 2 & 50 & 13 \\ -25 & -10 & -25 \\ 0 & 4 & 3 \end{bmatrix}$$

$5 \times 4 + 1 \times 2 + 7 \times 4 = 50$

Boolean matrix multiplication:

$$\begin{bmatrix} T & F & T \\ T & F & F \\ T & T & T \end{bmatrix} \times \begin{bmatrix} F & F & T \\ F & T & T \\ T & F & T \end{bmatrix} = \begin{bmatrix} & & \\ & & \\ & & \end{bmatrix}$$

Boolean Matrix Multiplication

Integer matrix multiplication:

$$\begin{bmatrix} 5 & 1 & 7 \\ -10 & 5 & 5 \\ 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} 1 & 4 & 0 \\ -3 & 2 & -8 \\ 0 & 4 & 3 \end{bmatrix} = \begin{bmatrix} 2 & 50 & 13 \\ -25 & -10 & -25 \\ 0 & 4 & 3 \end{bmatrix}$$

$5 \times 4 + 1 \times 2 + 7 \times 4 = 50$

Boolean matrix multiplication:

$$\begin{bmatrix} T & F & T \\ T & F & F \\ T & T & T \end{bmatrix} \times \begin{bmatrix} F & F & T \\ F & T & T \\ T & F & T \end{bmatrix} = \begin{bmatrix} & & \\ & & \\ & & \end{bmatrix}$$

Boolean Matrix Multiplication

Integer matrix multiplication:

$$\begin{bmatrix} 5 & 1 & 7 \\ -10 & 5 & 5 \\ 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} 1 & 4 & 0 \\ -3 & 2 & -8 \\ 0 & 4 & 3 \end{bmatrix} = \begin{bmatrix} 2 & 50 & 13 \\ -25 & -10 & -25 \\ 0 & 4 & 3 \end{bmatrix}$$

$5 \times 4 + 1 \times 2 + 7 \times 4 = 50$

Boolean matrix multiplication:

$$\begin{bmatrix} T & F & T \\ T & F & F \\ T & T & T \end{bmatrix} \times \begin{bmatrix} F & F & T \\ F & T & T \\ T & F & T \end{bmatrix} = \begin{bmatrix} \; & \; & \; \\ \; & \; & \; \\ \; & \; & \; \end{bmatrix}$$

Replace '$\times$' by '$\wedge$' and '$+$' by '$\vee$'

Boolean Matrix Multiplication

Integer matrix multiplication:

$$\begin{bmatrix} 5 & 1 & 7 \\ -10 & 5 & 5 \\ 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} 1 & 4 & 0 \\ -3 & 2 & -8 \\ 0 & 4 & 3 \end{bmatrix} = \begin{bmatrix} 2 & 50 & 13 \\ -25 & -10 & -25 \\ 0 & 4 & 3 \end{bmatrix}$$

$5 \times 4 + 1 \times 2 + 7 \times 4 = 50$

Boolean matrix multiplication:

$$\begin{bmatrix} T & F & T \\ T & F & F \\ T & T & T \end{bmatrix} \times \begin{bmatrix} F & F & T \\ F & T & T \\ T & F & T \end{bmatrix} = \begin{bmatrix} & & \\ & & \\ & & \end{bmatrix}$$

$(T \wedge F) \vee (F \wedge F) \vee (T \wedge T) = T$

Replace '$\times$' by '$\wedge$' and '$+$' by '$\vee$'

Boolean Matrix Multiplication

Integer matrix multiplication:

$$\begin{bmatrix} 5 & 1 & 7 \\ -10 & 5 & 5 \\ 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} 1 & 4 & 0 \\ -3 & 2 & -8 \\ 0 & 4 & 3 \end{bmatrix} = \begin{bmatrix} 2 & 50 & 13 \\ -25 & -10 & -25 \\ 0 & 4 & 3 \end{bmatrix}$$

$5 \times 4 + 1 \times 2 + 7 \times 4 = 50$

Boolean matrix multiplication:

$$\begin{bmatrix} T & F & T \\ T & F & F \\ T & T & T \end{bmatrix} \times \begin{bmatrix} F & F & T \\ F & T & T \\ T & F & T \end{bmatrix} = \begin{bmatrix} T & & \\ & & \\ & & \end{bmatrix}$$

$(T \wedge F) \vee (F \wedge F) \vee (T \wedge T) = T$

Replace '$\times$' by '$\wedge$' and '$+$' by '$\vee$'

Boolean Matrix Multiplication

Integer matrix multiplication:

$$\begin{bmatrix} 5 & 1 & 7 \\ -10 & 5 & 5 \\ 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} 1 & 4 & 0 \\ -3 & 2 & -8 \\ 0 & 4 & 3 \end{bmatrix} = \begin{bmatrix} 2 & 50 & 13 \\ -25 & -10 & -25 \\ 0 & 4 & 3 \end{bmatrix}$$

$5 \times 4 + 1 \times 2 + 7 \times 4 = 50$

Boolean matrix multiplication:

$$\begin{bmatrix} T & F & T \\ T & F & F \\ T & T & T \end{bmatrix} \times \begin{bmatrix} F & F & T \\ F & T & T \\ T & F & T \end{bmatrix} = \begin{bmatrix} T & \square & \\ & & \\ & & \end{bmatrix}$$

Replace '$\times$' by '$\wedge$' and '$+$' by '$\vee$'

Boolean Matrix Multiplication

Integer matrix multiplication:

$$\begin{bmatrix} 5 & 1 & 7 \\ -10 & 5 & 5 \\ 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} 1 & 4 & 0 \\ -3 & 2 & -8 \\ 0 & 4 & 3 \end{bmatrix} = \begin{bmatrix} 2 & 50 & 13 \\ -25 & -10 & -25 \\ 0 & 4 & 3 \end{bmatrix}$$

$5 \times 4 + 1 \times 2 + 7 \times 4 = 50$

Boolean matrix multiplication:

$$\begin{bmatrix} T & F & T \\ T & F & F \\ T & T & T \end{bmatrix} \times \begin{bmatrix} F & F & T \\ F & T & T \\ T & F & T \end{bmatrix} = \begin{bmatrix} T & \square & \\ & & \\ & & \end{bmatrix}$$

$(T \wedge F) \vee (F \wedge T) \vee (T \wedge F) = F$

Replace '$\times$' by '$\wedge$' and '$+$' by '$\vee$'

Boolean Matrix Multiplication

Integer matrix multiplication:

$$\begin{bmatrix} 5 & 1 & 7 \\ -10 & 5 & 5 \\ 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} 1 & 4 & 0 \\ -3 & 2 & -8 \\ 0 & 4 & 3 \end{bmatrix} = \begin{bmatrix} 2 & 50 & 13 \\ -25 & -10 & -25 \\ 0 & 4 & 3 \end{bmatrix}$$

$5 \times 4 + 1 \times 2 + 7 \times 4 = 50$

Boolean matrix multiplication:

$$\begin{bmatrix} T & F & T \\ T & F & F \\ T & T & T \end{bmatrix} \times \begin{bmatrix} F & F & T \\ F & T & T \\ T & F & T \end{bmatrix} = \begin{bmatrix} T & F & \\ & & \\ & & \end{bmatrix}$$

$(T \wedge F) \vee (F \wedge T) \vee (T \wedge F) = F$

Replace '$\times$' by '$\wedge$' and '$+$' by '$\vee$'

Boolean Matrix Multiplication

Integer matrix multiplication:

$$\begin{bmatrix} 5 & 1 & 7 \\ -10 & 5 & 5 \\ 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} 1 & 4 & 0 \\ -3 & 2 & -8 \\ 0 & 4 & 3 \end{bmatrix} = \begin{bmatrix} 2 & 50 & 13 \\ -25 & -10 & -25 \\ 0 & 4 & 3 \end{bmatrix}$$

$5 \times 4 + 1 \times 2 + 7 \times 4 = 50$

Boolean matrix multiplication:

$$\begin{bmatrix} T & F & T \\ T & F & F \\ T & T & T \end{bmatrix} \times \begin{bmatrix} F & F & T \\ F & T & T \\ T & F & T \end{bmatrix} = \begin{bmatrix} T & F & T \\ F & F & T \\ T & T & T \end{bmatrix}$$

$(T \wedge F) \vee (F \wedge T) \vee (T \wedge F) = F$

Replace '$\times$' by '$\wedge$' and '$+$' by '$\vee$'

Boolean Matrix Multiplication and Set Intersection

$$\overset{A}{\begin{bmatrix} T & F & T \\ T & F & F \\ T & T & T \end{bmatrix}} \times \overset{B}{\begin{bmatrix} F & F & T \\ F & T & T \\ T & F & T \end{bmatrix}} = \overset{C}{\begin{bmatrix} T & F & T \\ F & F & T \\ T & T & T \end{bmatrix}}$$

Boolean Matrix Multiplication and Set Intersection

$$\overset{A}{\begin{bmatrix} T & F & T \\ T & F & F \\ T & T & T \end{bmatrix}} \times \overset{B}{\begin{bmatrix} F & F & T \\ F & T & T \\ T & F & T \end{bmatrix}} = \overset{C}{\begin{bmatrix} T & F & T \\ F & F & T \\ T & T & T \end{bmatrix}}$$

$C_{i,j} = T$ if and only if there exists an ℓ s.t. $A_{i,\ell} = B_{\ell,j} = T$

Boolean Matrix Multiplication and Set Intersection

$$\overset{A}{\begin{bmatrix} T & F & T \\ T & F & F \\ T & T & T \end{bmatrix}} \times \overset{B}{\begin{bmatrix} F & F & T \\ F & T & T \\ T & F & T \end{bmatrix}} = \overset{C}{\begin{bmatrix} T & F & T \\ F & F & T \\ T & T & T \end{bmatrix}}$$

$C_{i,j} = T$ if and only if there exists an ℓ s.t. $A_{i,\ell} = B_{\ell,j} = T$

Eg. $C_{2,3} = T$, $A_{2,1} = B_{1,3} = T$, so $\ell = 1$

Boolean Matrix Multiplication and Set Intersection

$$\overset{A}{\begin{bmatrix} T & F & T \\ T & F & F \\ T & T & T \end{bmatrix}} \times \overset{B}{\begin{bmatrix} F & F & T \\ F & T & T \\ T & F & T \end{bmatrix}} = \overset{C}{\begin{bmatrix} T & F & T \\ F & F & T \\ T & T & T \end{bmatrix}}$$

$C_{i,j} = T$ if and only if there exists an ℓ s.t. $A_{i,\ell} = B_{\ell,j} = T$

Eg. $C_{2,3} = T$, $A_{2,1} = B_{1,3} = T$, so $\ell = 1$

Let $A_i = \{j \mid A_{i,j} = T\}$ and $B_j = \{i \mid B_{i,j} = T\}$

Boolean Matrix Multiplication and Set Intersection

$$\overset{A}{\begin{bmatrix} T & F & T \\ \boxed{T} & F & F \\ T & T & T \end{bmatrix}} \times \overset{B}{\begin{bmatrix} F & F & \boxed{T} \\ F & T & T \\ T & F & T \end{bmatrix}} = \overset{C}{\begin{bmatrix} T & F & T \\ F & F & \boxed{T} \\ T & T & T \end{bmatrix}}$$

$C_{i,j} = T$ if and only if there exists an ℓ s.t. $A_{i,\ell} = B_{\ell,j} = T$

Eg. $C_{2,3} = T$, $A_{2,1} = B_{1,3} = T$, so $\ell = 1$

Let $A_i = \{j \mid A_{i,j} = T\}$ and $B_j = \{i \mid B_{i,j} = T\}$

Then $C_{i,j} = T$ if and only if $A_i \cap B_j \neq \emptyset$

Boolean Matrix Multiplication and Set Intersection

$$\overset{A}{\begin{bmatrix} T & F & T \\ T & F & F \\ T & T & T \end{bmatrix}} \times \overset{B}{\begin{bmatrix} F & F & T \\ F & T & T \\ T & F & T \end{bmatrix}} = \overset{C}{\begin{bmatrix} T & F & T \\ F & F & T \\ T & T & T \end{bmatrix}}$$

$C_{i,j} = T$ if and only if there exists an ℓ s.t. $A_{i,\ell} = B_{\ell,j} = T$

Eg. $C_{2,3} = T$, $A_{2,1} = B_{1,3} = T$, so $\ell = 1$

Let $A_i = \{j \mid A_{i,j} = T\}$ and $B_j = \{i \mid B_{i,j} = T\}$

Then $C_{i,j} = T$ if and only if $A_i \cap B_j \neq \emptyset$

Boolean Matrix Multiplication and Set Intersection

$$\overset{A}{\begin{bmatrix} T & F & T \\ T & F & F \\ T & T & T \end{bmatrix}} \times \overset{B}{\begin{bmatrix} F & F & T \\ F & T & T \\ T & F & T \end{bmatrix}} = \overset{C}{\begin{bmatrix} T & F & T \\ F & F & T \\ T & T & T \end{bmatrix}}$$

$C_{i,j} = T$ if and only if there exists an ℓ s.t. $A_{i,\ell} = B_{\ell,j} = T$

Eg. $C_{2,3} = T$, $A_{2,1} = B_{1,3} = T$, so $\ell = 1$

Let $A_i = \{j \mid A_{i,j} = T\}$ and $B_j = \{i \mid B_{i,j} = T\}$

Then $C_{i,j} = T$ if and only if $A_i \cap B_j \neq \emptyset$

Eg. $A_2 = \{1\}$, $B_3 = \{1, 2, 3\}$. $A_2 \cap B_3 = \{1\} \neq \emptyset \Rightarrow C_{2,3} = 1$

Two Patterns - Simple Solution

Two Patterns - Simple Solution

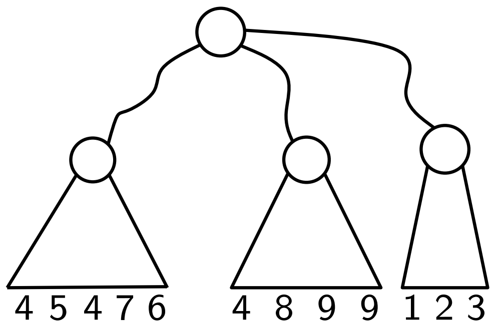
Build suffix tree on concatenation of the documents

$$S = d_1\$d_2\$\cdots\$d_D\$$$

Two Patterns - Simple Solution

Build suffix tree on concatenation of the documents

$$S = d_1\$d_2\$\cdots\$d_D\$$$

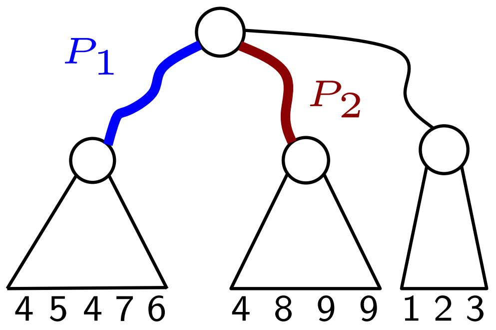


Annotate each leaf with the document id the suffx starts in

Two Patterns - Simple Solution

Build suffix tree on concatenation of the documents

$$S = d_1\$d_2\$\cdots\$d_D\$$$



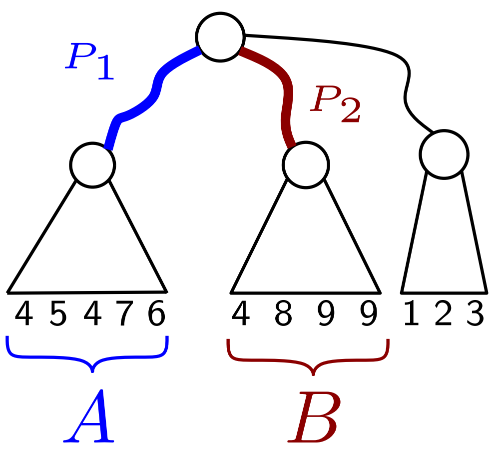
Annotate each leaf with the document id the suffx starts in

Query: P_1 and P_2

Two Patterns - Simple Solution

Build suffix tree on concatenation of the documents

$$S = d_1\$d_2\$\cdots\$d_D\$$$



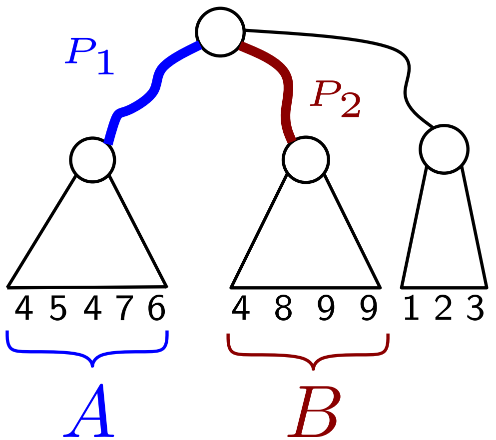
Annotate each leaf with the document id the suffx starts in

Query: P_1 and P_2

Two Patterns - Simple Solution

Build suffix tree on concatenation of the documents

$$S = d_1\$d_2\$\cdots\$d_D\$$$



Annotate each leaf with the document id the suffx starts in

Query: P_1 and P_2

Report: $A \cap B$

Two Patterns - Reduction from Matrix Multiplication

Two Patterns - Reduction from Matrix Multiplication

$$\overset{A}{\begin{bmatrix} T & F & T \\ T & F & F \\ T & T & T \end{bmatrix}} \times \overset{B}{\begin{bmatrix} F & F & T \\ F & T & T \\ T & F & T \end{bmatrix}} = \overset{C}{\begin{bmatrix} & & \\ & & \\ & & \end{bmatrix}}$$

$C_{i,j} = T \Leftrightarrow \exists \ell$ s.t.
$A_{i,\ell} = B_{\ell,j} = T$

Two Patterns - Reduction from Matrix Multiplication

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array} \overset{A}{\begin{bmatrix} T & F & T \\ T & F & F \\ T & T & T \end{bmatrix}} \times \overset{B}{\begin{bmatrix} F & F & T \\ F & T & T \\ T & F & T \end{bmatrix}} = \overset{C}{\begin{bmatrix} & & \\ & & \\ & & \end{bmatrix}}$$

$C_{i,j} = T \Leftrightarrow \exists \ell$ s.t. $A_{i,\ell} = B_{\ell,j} = T$

$$d_1^A = 1\#2\#3\#$$

Two Patterns - Reduction from Matrix Multiplication

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array} \overset{A}{\begin{bmatrix} T & F & T \\ T & F & F \\ T & T & T \end{bmatrix}} \times \overset{B}{\begin{bmatrix} F & F & T \\ F & T & T \\ T & F & T \end{bmatrix}} = \overset{C}{\begin{bmatrix} & & \\ & & \\ & & \end{bmatrix}}$$

$C_{i,j} = T \Leftrightarrow \exists \ell$ s.t. $A_{i,\ell} = B_{\ell,j} = T$

$$d_1^A = 1\#2\#3\#$$

$$d_2^A = 3\#$$

Two Patterns - Reduction from Matrix Multiplication

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array} \overset{A}{\begin{bmatrix} T & F & T \\ T & F & F \\ T & T & T \end{bmatrix}} \times \overset{B}{\begin{bmatrix} F & F & T \\ F & T & T \\ T & F & T \end{bmatrix}} = \overset{C}{\begin{bmatrix} & & \\ & & \\ & & \end{bmatrix}}$$

$C_{i,j} = T \Leftrightarrow \exists \ell$ s.t. $A_{i,\ell} = B_{\ell,j} = T$

$$d_1^A = 1\#2\#3\#$$

$$d_2^A = 3\#$$

$$d_3^A = 1\#3\#$$

Two Patterns - Reduction from Matrix Multiplication

$$\begin{matrix} 1 \\ 2 \\ 3 \end{matrix} \overset{A}{\begin{bmatrix} T & F & T \\ T & F & F \\ T & T & T \end{bmatrix}} \times \overset{B}{\underset{\begin{matrix} 4 & 5 & 6 \end{matrix}}{\begin{bmatrix} \boxed{F \quad F \quad T} \\ F \quad T \quad T \\ T \quad F \quad T \end{bmatrix}}} = \overset{C}{\begin{bmatrix} & & \\ & & \\ & & \end{bmatrix}}$$

$C_{i,j} = T \Leftrightarrow \exists \ell$ s.t. $A_{i,\ell} = B_{\ell,j} = T$

$$d_1^A = 1\#2\#3\# \qquad d_1^B = 6\#$$

$$d_2^A = 3\#$$

$$d_3^A = 1\#3\#$$

Two Patterns - Reduction from Matrix Multiplication

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array} \overset{A}{\begin{bmatrix} T & F & T \\ T & F & F \\ T & T & T \end{bmatrix}} \times \overset{B}{\begin{bmatrix} F & F & T \\ \boxed{F \quad T \quad T} \\ T & F & T \end{bmatrix}} = \overset{C}{\begin{bmatrix} & & \\ & & \\ & & \end{bmatrix}}$$

$$\begin{array}{ccc} 4 & 5 & 6 \end{array}$$

$C_{i,j} = T \Leftrightarrow \exists \ell$ s.t. $A_{i,\ell} = B_{\ell,j} = T$

$$d_1^A = 1\#2\#3\# \qquad d_1^B = 6\#$$

$$d_2^A = 3\# \qquad d_2^B = 5\#6\#$$

$$d_3^A = 1\#3\#$$

Two Patterns - Reduction from Matrix Multiplication

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array} \overset{A}{\begin{bmatrix} T & F & T \\ T & F & F \\ T & T & T \end{bmatrix}} \times \overset{B}{\underset{4 \quad 5 \quad 6}{\begin{bmatrix} F & F & T \\ F & T & T \\ \boxed{T \quad F \quad T} \end{bmatrix}}} = \overset{C}{\begin{bmatrix} & & \\ & & \\ & & \end{bmatrix}}$$

$C_{i,j} = T \Leftrightarrow \exists \ell$ s.t. $A_{i,\ell} = B_{\ell,j} = T$

$$d_1^A = 1\#2\#3\# \qquad d_1^B = 6\#$$

$$d_2^A = 3\# \qquad d_2^B = 5\#6\#$$

$$d_3^A = 1\#3\# \qquad d_3^B = 4\#6\#$$

Two Patterns - Reduction from Matrix Multiplication

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array} \overset{A}{\begin{bmatrix} T & F & T \\ T & F & F \\ T & T & T \end{bmatrix}} \times \overset{B}{\underset{\begin{matrix} 4 & 5 & 6 \end{matrix}}{\begin{bmatrix} F & F & T \\ F & T & T \\ T & F & T \end{bmatrix}}} = \overset{C}{\begin{bmatrix} & & \\ & & \\ & & \end{bmatrix}}$$

$C_{i,j} = T \Leftrightarrow \exists \ell$ s.t. $A_{i,\ell} = B_{\ell,j} = T$

$d_1^A = 1\#2\#3\#$ $\qquad d_1^B = 6\#$ $\qquad d_1 = 1\#2\#3\#6\#$

$d_2^A = 3\#$ $\qquad d_2^B = 5\#6\#$ $\qquad d_2 = 3\#5\#6\#$

$d_3^A = 1\#3\#$ $\qquad d_3^B = 4\#6\#$ $\qquad d_3 = 1\#3\#4\#6\#$

Two Patterns - Reduction from Matrix Multiplication

$$\begin{matrix} & A & & B & & C \\ \begin{matrix}1\\2\\3\end{matrix} & \begin{bmatrix} T & F & T \\ T & F & F \\ T & T & T \end{bmatrix} & \times & \begin{bmatrix} F & F & T \\ F & T & T \\ T & F & T \end{bmatrix} & = & \begin{bmatrix} & & \\ & & \\ & & \end{bmatrix} \\ & & & \begin{matrix}4 & 5 & 6\end{matrix} & & \end{matrix}$$

$C_{i,j} = T \Leftrightarrow \exists \ell$ s.t. $A_{i,\ell} = B_{\ell,j} = T$

$d_1^A = 1\#2\#3\#$ $\quad d_1^B = 6\#$ $\quad d_1 = 1\#2\#3\#6\#$

$d_2^A = 3\#$ $\quad d_2^B = 5\#6\#$ $\quad d_2 = 3\#5\#6\#$

$d_3^A = 1\#3\#$ $\quad d_3^B = 4\#6\#$ $\quad d_3 = 1\#3\#4\#6\#$

D

Two Patterns - Reduction from Matrix Multiplication

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array} \overset{A}{\begin{bmatrix} T & F & T \\ T & F & F \\ T & T & T \end{bmatrix}} \times \overset{B}{\begin{bmatrix} F & F & T \\ F & T & T \\ T & F & T \end{bmatrix}} = \overset{C}{\begin{bmatrix} & & \\ & & \\ & & \end{bmatrix}}$$

$$\begin{array}{ccc} 4 & 5 & 6 \end{array}$$

$C_{i,j} = T \Leftrightarrow \exists \ell$ s.t. $A_{i,\ell} = B_{\ell,j} = T$

$d_1^A = 1\#2\#3\#$ $\quad d_1^B = 6\#$ $\quad d_1 = 1\#2\#3\#6\#$

$d_2^A = 3\#$ $\quad d_2^B = 5\#6\#$ $\quad d_2 = 3\#5\#6\#$

$d_3^A = 1\#3\#$ $\quad d_3^B = 4\#6\#$ $\quad d_3 = 1\#3\#4\#6\#$

D

Query: $P_1 =$ 1 $P_2 =$ 4

Two Patterns - Reduction from Matrix Multiplication

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array} \overset{A}{\begin{bmatrix} T & F & T \\ T & F & F \\ T & T & T \end{bmatrix}} \times \overset{B}{\underset{4 \quad 5 \quad 6}{\begin{bmatrix} F & F & T \\ F & T & T \\ T & F & T \end{bmatrix}}} = \overset{C}{\begin{bmatrix} T & & \\ & & \\ & & \end{bmatrix}}$$

$C_{i,j} = T \Leftrightarrow \exists \ell$ s.t. $A_{i,\ell} = B_{\ell,j} = T$

$d_1^A = 1\#2\#3\#$ $\quad d_1^B = 6\#$ $\quad d_1 = 1\#2\#3\#6\#$

$d_2^A = 3\#$ $\quad d_2^B = 5\#6\#$ $\quad d_2 = 3\#5\#6\#$

$d_3^A = 1\#3\#$ $\quad d_3^B = 4\#6\#$ $\quad d_3 = 1\#3\#4\#6\#$ $\quad D$

Query: $P_1 =$ 1 $\quad P_2 =$ 4

Two Patterns - Reduction from Matrix Multiplication

$$\begin{matrix} 1 \\ 2 \\ 3 \end{matrix} \overset{A}{\begin{bmatrix} T & F & T \\ T & F & F \\ T & T & T \end{bmatrix}} \times \overset{B}{\begin{bmatrix} F & F & T \\ F & T & T \\ T & F & T \end{bmatrix}} = \overset{C}{\begin{bmatrix} T & & \\ & & \\ & & \end{bmatrix}}$$

$$\begin{matrix} 4 & 5 & 6 \end{matrix}$$

$C_{i,j} = T \Leftrightarrow \exists \ell$ s.t. $A_{i,\ell} = B_{\ell,j} = T$

$$d_1^A = 1\#2\#3\# \qquad d_1^B = 6\# \qquad d_1 = 1\#2\#3\#6\#$$

$$d_2^A = 3\# \qquad d_2^B = 5\#6\# \qquad d_2 = 3\#5\#6\#$$

$$d_3^A = 1\#3\# \qquad d_3^B = 4\#6\# \qquad d_3 = 1\#3\#4\#6\#$$

D

Query: $P_1 =$ 1 $P_2 =$ 5

Two Patterns - Reduction from Matrix Multiplication

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array} \overset{A}{\begin{bmatrix} T & F & T \\ T & F & F \\ T & T & T \end{bmatrix}} \times \underset{\begin{matrix} 4 & 5 & 6 \end{matrix}}{\overset{B}{\begin{bmatrix} F & F & T \\ F & T & T \\ T & F & T \end{bmatrix}}} = \overset{C}{\begin{bmatrix} T & F & \\ & & \\ & & \end{bmatrix}}$$

$C_{i,j} = T \Leftrightarrow \exists \ell$ s.t. $A_{i,\ell} = B_{\ell,j} = T$

$d_1^A = 1\#2\#3\#$ | $d_1^B = 6\#$ | $d_1 = 1\#2\#3\#6\#$

$d_2^A = 3\#$ | $d_2^B = 5\#6\#$ | $d_2 = 3\#5\#6\#$

$d_3^A = 1\#3\#$ | $d_3^B = 4\#6\#$ | $d_3 = 1\#3\#4\#6\#$ D

Query: $P_1 =$ 1 $P_2 =$ 5

$i \in d_\ell \Leftrightarrow A_{i,\ell} = T$

$j + 3 \in d_\ell \Leftrightarrow B_{\ell,j} = T$

Two Patterns - Reduction from Matrix Multiplication

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array} \overset{A}{\begin{bmatrix} T & F & T \\ T & F & F \\ T & T & T \end{bmatrix}} \times \underset{4 \quad 5 \quad 6}{\overset{B}{\begin{bmatrix} F & F & T \\ F & T & T \\ T & F & T \end{bmatrix}}} = \overset{C}{\begin{bmatrix} T & F & T \\ F & F & T \\ T & & \end{bmatrix}}$$

$C_{i,j} = T \Leftrightarrow \exists \ell$ s.t. $A_{i,\ell} = B_{\ell,j} = T$

$$d_1^A = 1\#2\#3\# \qquad d_1^B = 6\# \qquad d_1 = 1\#2\#3\#6\#$$

$$d_2^A = 3\# \qquad d_2^B = 5\#6\# \qquad d_2 = 3\#5\#6\#$$

$$d_3^A = 1\#3\# \qquad d_3^B = 4\#6\# \qquad d_3 = 1\#3\#4\#6\#$$

D

$i \in d_\ell \Leftrightarrow A_{i,\ell} = T$

$j + 3 \in d_\ell \Leftrightarrow B_{\ell,j} = T$

Two Patterns - Reduction from Matrix Multiplication

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array} \overset{A}{\begin{bmatrix} T & F & T \\ T & F & F \\ T & T & T \end{bmatrix}} \times \overset{B}{\underset{\begin{array}{ccc} 4 & 5 & 6 \end{array}}{\begin{bmatrix} F & F & T \\ F & T & T \\ T & F & T \end{bmatrix}}} = \overset{C}{\begin{bmatrix} T & F & T \\ F & F & T \\ T & & \end{bmatrix}}$$

$C_{i,j} = T \Leftrightarrow \exists \ell$ s.t. $A_{i,\ell} = B_{\ell,j} = T$

$$d_1^A = 1\#2\#3\# \qquad d_1^B = 6\# \qquad d_1 = 1\#2\#3\#6\#$$

$$d_2^A = 3\# \qquad d_2^B = 5\#6\# \qquad d_2 = 3\#5\#6\#$$

$$d_3^A = 1\#3\# \qquad d_3^B = 4\#6\# \qquad d_3 = 1\#3\#4\#6\#$$

D

Query: $P_1 =$ 3 $P_2 =$ 5

$i \in d_\ell \Leftrightarrow A_{i,\ell} = T$

$j + 3 \in d_\ell \Leftrightarrow B_{\ell,j} = T$

Two Patterns - Reduction from Matrix Multiplication

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array} \overset{A}{\begin{bmatrix} T & F & T \\ T & F & F \\ T & T & T \end{bmatrix}} \times \underset{\;4\quad\;5\quad\;6}{\overset{B}{\begin{bmatrix} F & F & T \\ F & T & T \\ T & F & T \end{bmatrix}}} = \overset{C}{\begin{bmatrix} T & F & T \\ F & F & T \\ T & T & \end{bmatrix}}$$

$C_{i,j} = T \Leftrightarrow \exists \ell$ s.t. $A_{i,\ell} = B_{\ell,j} = T$

$d_1^A = 1\#2\#3\#$ $\quad d_1^B = 6\#$

$d_2^A = 3\#$ $\quad d_2^B = 5\#6\#$

$d_3^A = 1\#3\#$ $\quad d_3^B = 4\#6\#$

D:

$d_1 = 1\#2\#3\#6\#$

$d_2 = 3\#5\#6\#$

$d_3 = 1\#3\#4\#6\#$

Query: $P_1 =$ 3 $\quad P_2 =$ 5

$i \in d_\ell \Leftrightarrow A_{i,\ell} = T$

$j + 3 \in d_\ell \Leftrightarrow B_{\ell,j} = T$

Two Patterns - Reduction from Matrix Multiplication

$$\begin{matrix} & A \\ \begin{matrix}1\\2\\3\end{matrix} & \begin{bmatrix} T & F & T \\ T & F & F \\ T & T & T \end{bmatrix} \end{matrix} \times \begin{matrix} B \\ \begin{bmatrix} F & F & T \\ F & T & T \\ T & F & T \end{bmatrix} \\ \begin{matrix}4 & 5 & 6\end{matrix} \end{matrix} = \begin{matrix} C \\ \begin{bmatrix} T & F & T \\ F & F & T \\ T & T & \end{bmatrix} \end{matrix}$$

$C_{i,j} = T \Leftrightarrow \exists \ell$ s.t. $A_{i,\ell} = B_{\ell,j} = T$

$$d_1^A = 1\#2\#3\# \qquad d_1^B = 6\# \qquad d_1 = 1\#2\#3\#6\#$$

$$d_2^A = 3\# \qquad d_2^B = 5\#6\# \qquad d_2 = 3\#5\#6\#$$

$$d_3^A = 1\#3\# \qquad d_3^B = 4\#6\# \qquad d_3 = 1\#3\#4\#6\#$$

D

Query: $P_1 =$ 3 $P_2 =$ 6

$i \in d_\ell \Leftrightarrow A_{i,\ell} = T$

$j + 3 \in d_\ell \Leftrightarrow B_{\ell,j} = T$

Two Patterns - Reduction from Matrix Multiplication

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array} \overset{A}{\begin{bmatrix} T & F & T \\ T & F & F \\ T & T & T \end{bmatrix}} \times \overset{B}{\underset{\begin{array}{ccc} 4 & 5 & 6 \end{array}}{\begin{bmatrix} F & F & T \\ F & T & T \\ T & F & T \end{bmatrix}}} = \overset{C}{\begin{bmatrix} T & F & T \\ F & F & T \\ T & T & T \end{bmatrix}}$$

$C_{i,j} = T \Leftrightarrow \exists \ell$ s.t. $A_{i,\ell} = B_{\ell,j} = T$

$d_1^A = 1\#2\#3\#$ $\quad d_1^B = 6\#$ $\quad d_1 = 1\#2\#3\#6\#$

$d_2^A = 3\#$ $\quad d_2^B = 5\#6\#$ $\quad d_2 = 3\#5\#6\#$

$d_3^A = 1\#3\#$ $\quad d_3^B = 4\#6\#$ $\quad d_3 = 1\#3\#4\#6\#$ $\quad D$

Query: $P_1 =$ 3 $\quad P_2 =$ 6

$i \in d_\ell \Leftrightarrow A_{i,\ell} = T$

$j + 3 \in d_\ell \Leftrightarrow B_{\ell,j} = T$

Summary

Summary

Current best matrix multiplication algorithm:

Summary

Current best matrix multiplication algorithm:

$$O(\sqrt{n}^{2.3728639}) = O(n^{1.18643195}) \text{ (algebraic)}$$

Summary

Current best matrix multiplication algorithm:

$$O(\sqrt{n}^{2.3728639}) = O(n^{1.18643195}) \text{ (algebraic)}$$

Reduction: Construction + n queries solve matrix multiplication

Summary

Current best matrix multiplication algorithm:

$$O(\sqrt{n}^{2.3728639}) = O(n^{1.18643195}) \text{ (algebraic)}$$

Reduction: Construction $+\ n$ queries solve matrix multiplication

$\Rightarrow$ either $\tilde{\Omega}(n^{1.18})$ construction time or

Summary

Current best matrix multiplication algorithm:

$$O(\sqrt{n}^{2.3728639}) = O(n^{1.18643195}) \text{ (algebraic)}$$

Reduction: Construction $+ n$ queries solve matrix multiplication

$\Rightarrow$ either $\tilde{\Omega}(n^{1.18})$ construction time or

$$\tilde{\Omega}(n^{1.18}/n) = \tilde{\Omega}(n^{0.18}) \text{ time per query}$$

Summary

Current best matrix multiplication algorithm:

$$O(\sqrt{n}^{2.3728639}) = O(n^{1.18643195}) \text{ (algebraic)}$$

Reduction: Construction $+\ n$ queries solve matrix multiplication

$\Rightarrow$ either $\tilde{\Omega}(n^{1.18})$ construction time or

$\tilde{\Omega}(n^{1.18}/n) = \tilde{\Omega}(n^{0.18})$ time per query

(we lose a log factor in the reduction)

Summary

Current best matrix multiplication algorithm:

$$O(\sqrt{n}^{2.3728639}) = O(n^{1.18643195}) \text{ (algebraic)}$$

$$\tilde{O}(\sqrt{n}^{3}) = \tilde{O}(n^{1.5}) \text{ (combinatorial)}$$

Reduction: Construction + n queries solve matrix multiplication

$\Rightarrow$ either $\tilde{\Omega}(n^{1.18})$ construction time or

$\tilde{\Omega}(n^{1.18}/n) = \tilde{\Omega}(n^{0.18})$ time per query

(we lose a log factor in the reduction)

Summary

Current best matrix multiplication algorithm:

$$O(\sqrt{n}^{2.3728639}) = O(n^{1.18643195}) \text{ (algebraic)}$$

$$\tilde{O}(\sqrt{n}^{3}) = \tilde{O}(n^{1.5}) \text{ (combinatorial)}$$

Reduction: Construction + n queries solve matrix multiplication

$\Rightarrow$ either $\tilde{\Omega}(n^{1.18})$ construction time or $\tilde{\Omega}(n^{1.18}/n) = \tilde{\Omega}(n^{0.18})$ time per query

$\tilde{\Omega}(n^{1.5})$ / $\tilde{\Omega}(n^{1.5}/n) = \tilde{\Omega}(\sqrt{n})$

(we lose a log factor in the reduction)

Summary

Current best matrix multiplication algorithm:

$$O(\sqrt{n}^{2.3728639}) = O(n^{1.18643195}) \text{ (algebraic)}$$

$$\tilde{O}(\sqrt{n}^{3}) = \tilde{O}(n^{1.5}) \text{ (combinatorial)}$$

Reduction: Construction $+\ n$ queries solve matrix multiplication

$\Rightarrow$ either $\tilde{\Omega}(n^{1.18})$ construction time or $\quad \tilde{\Omega}(n^{1.5})$

$\tilde{\Omega}(n^{1.18}/n) = \tilde{\Omega}(n^{0.18})$ time per query $\quad \tilde{\Omega}(n^{1.5}/n) = \tilde{\Omega}(\sqrt{n})$

(we lose a log factor in the reduction)

Recall Hon et. al:

Summary

Current best matrix multiplication algorithm:

$$O(\sqrt{n}^{2.3728639}) = O(n^{1.18643195}) \text{ (algebraic)}$$

$$\tilde{O}(\sqrt{n}^{3}) = \tilde{O}(n^{1.5}) \text{ (combinatorial)}$$

Reduction: Construction + n queries solve matrix multiplication

$\Rightarrow$ either $\tilde{\Omega}(n^{1.18})$ construction time or $\tilde{\Omega}(n^{1.18}/n) = \tilde{\Omega}(n^{0.18})$ time per query — $\tilde{\Omega}(n^{1.5})$ / $\tilde{\Omega}(n^{1.5}/n) = \tilde{\Omega}(\sqrt{n})$

(we lose a log factor in the reduction)

Recall Hon et. al: $O(n)$ space and $O(\sqrt{n})$ counting

Thank You

2-Dimensional Substring Indexing

2-Dimensional Substring Indexing

Index pairs of documents: $\{(d_{1,1}, d_{1,2}), (d_{2,1}, d_{2,2}), (d_{3,1}, d_{3,2}) \ldots\}$

such that we can report all pairs where P_1 occurs in the first and P_2 occurs in the second document

2-Dimensional Substring Indexing

Index pairs of documents: $\{(d_{1,1}, d_{1,2}), (d_{2,1}, d_{2,2}), (d_{3,1}, d_{3,2}) \ldots\}$

such that we can report all pairs where P_1 occurs in the first and P_2 occurs in the second document

Reduces to 'common colors problem'

2-Dimensional Substring Indexing

Index pairs of documents: $\{(d_{1,1}, d_{1,2}), (d_{2,1}, d_{2,2}), (d_{3,1}, d_{3,2}) \ldots\}$

such that we can report all pairs where P_1 occurs in the first and P_2 occurs in the second document

Reduces to 'common colors problem'

Index an array of colors, such that when given two intervals: report all colors common to both intervals.

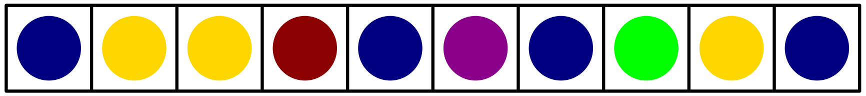
2-Dimensional Substring Indexing

Index pairs of documents: $\{(d_{1,1}, d_{1,2}), (d_{2,1}, d_{2,2}), (d_{3,1}, d_{3,2}) \ldots\}$

such that we can report all pairs where P_1 occurs in the first and P_2 occurs in the second document

Reduces to 'common colors problem'

Index an array of colors, such that when given two intervals: report all colors common to both intervals.



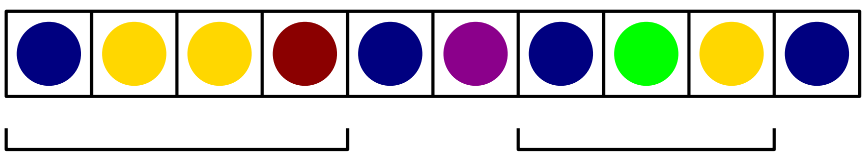
2-Dimensional Substring Indexing

Index pairs of documents: $\{(d_{1,1}, d_{1,2}), (d_{2,1}, d_{2,2}), (d_{3,1}, d_{3,2}) \ldots\}$

such that we can report all pairs where P_1 occurs in the first and P_2 occurs in the second document

Reduces to 'common colors problem'

Index an array of colors, such that when given two intervals: report all colors common to both intervals.



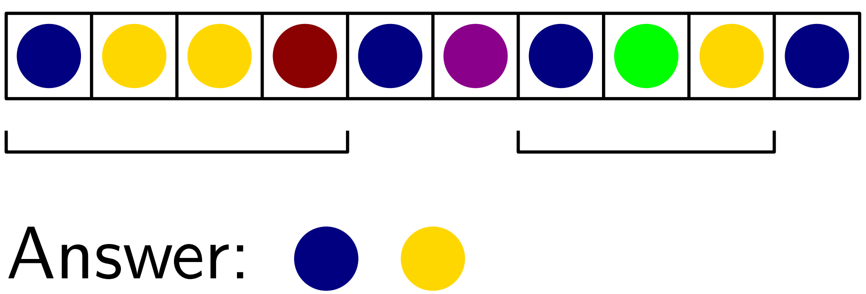
2-Dimensional Substring Indexing

Index pairs of documents: $\{(d_{1,1}, d_{1,2}), (d_{2,1}, d_{2,2}), (d_{3,1}, d_{3,2}) \ldots\}$

such that we can report all pairs where P_1 occurs in the first and P_2 occurs in the second document

Reduces to 'common colors problem'

Index an array of colors, such that when given two intervals: report all colors common to both intervals.

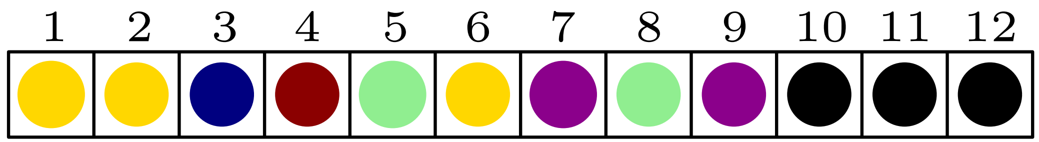


Common Colors - Warm Up

Given two intervals, are their colors disjoint?

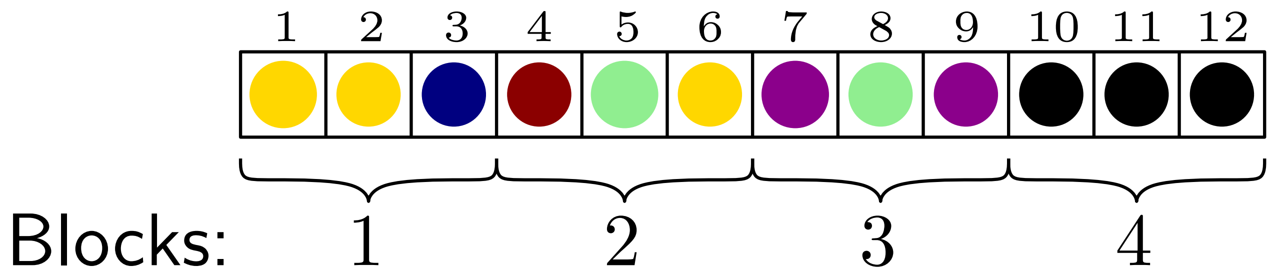
Common Colors - Warm Up

Given two intervals, are their colors disjoint?



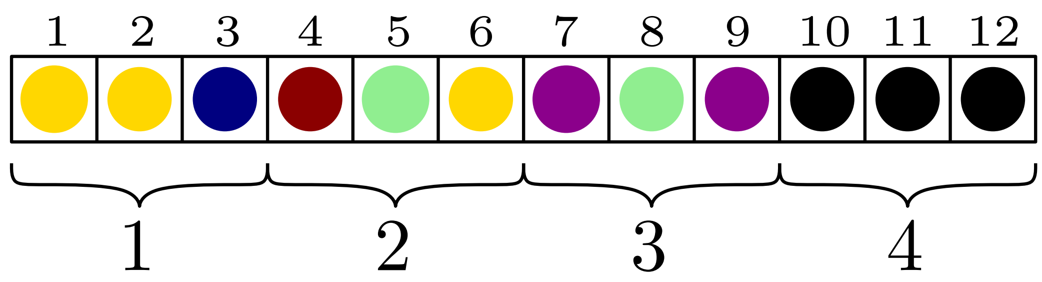
Common Colors - Warm Up

Given two intervals, are their colors disjoint?



Common Colors - Warm Up

Given two intervals, are their colors disjoint?

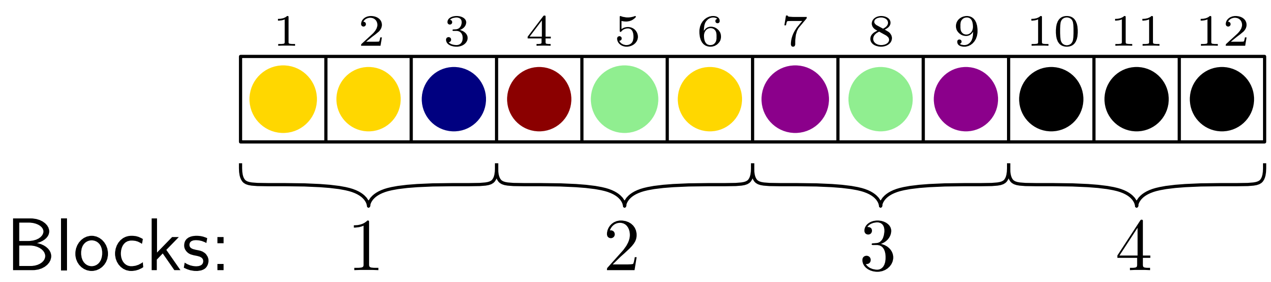


Blocks:

	1	2	3	4
1	1	1	0	0
2	1	1	1	0
3	0	1	1	0
4	0	0	0	1

Common Colors - Warm Up

Given two intervals, are their colors disjoint?

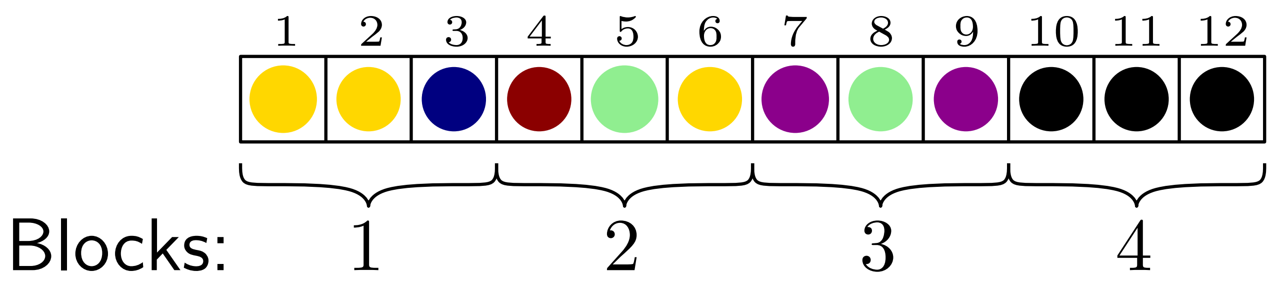


	1	2	3	4
1	1	1	0	0
2	1	1	1	0
3	0	1	1	0
4	0	0	0	1

Query: $[1, 7]$, $[9, 12]$

Common Colors - Warm Up

Given two intervals, are their colors disjoint?

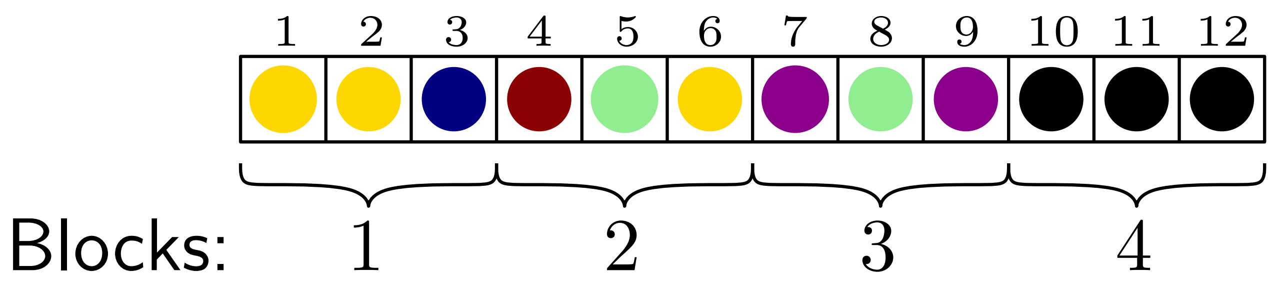


	1	2	3	4
1	1	1	0	0
2	1	1	1	0
3	0	1	1	0
4	0	0	0	1

Query: $[1, 7]$, $[9, 12]$

Common Colors - Warm Up

Given two intervals, are their colors disjoint?



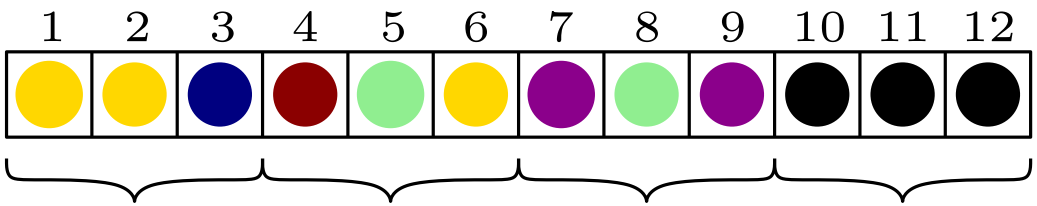
	1	2	3	4
1	1	1	0	0
2	1	1	1	0
3	0	1	1	0
4	0	0	0	1

Query: $[1, 7]$, $[9, 12]$

Check remaining indices

Common Colors - Warm Up

Given two intervals, are their colors disjoint?



Blocks: 1 2 3 4

	1	2	3	4
1	1	1	0	0
2	1	1	1	0
3	0	1	1	0
4	0	0	0	1

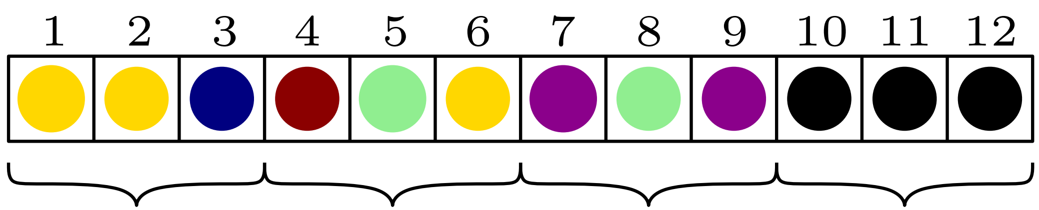
Query: $[1, 7]$, $[9, 12]$

Check remaining indices

 Occurs at 7 and 9

Common Colors - Warm Up

Given two intervals, are their colors disjoint?



Blocks: 1 2 3 4

	1	2	3	4
1	1	1	0	0
2	1	1	1	0
3	0	1	1	0
4	0	0	0	1

(RMQ)

Query: $[1, 7]$, $[9, 12]$

Check remaining indices

Occurs at 7 and 9

(1D range emptiness)

Common Colors - Reporting

Common Colors - Reporting

Idea: Build 'balanced' binary tree on colors

- Each node has an associated color
- Each node induces a subset of colors corresponding to the nodes in its subtree

Common Colors - Reporting

Idea: Build 'balanced' binary tree on colors

- Each node has an associated color
- Each node induces a subset of colors corresponding to the nodes in its subtree

Algorithm: Preorder traversal of the tree

Common Colors - Reporting

Idea: Build 'balanced' binary tree on colors

Each node has an associated color

Each node induces a subset of colors corresponding to the nodes in its subtree

Algorithm: Preorder traversal of the tree

2 queries at node u:

Common Colors - Reporting

Idea: Build 'balanced' binary tree on colors

Each node has an associated color

Each node induces a subset of colors corresponding to the nodes in its subtree

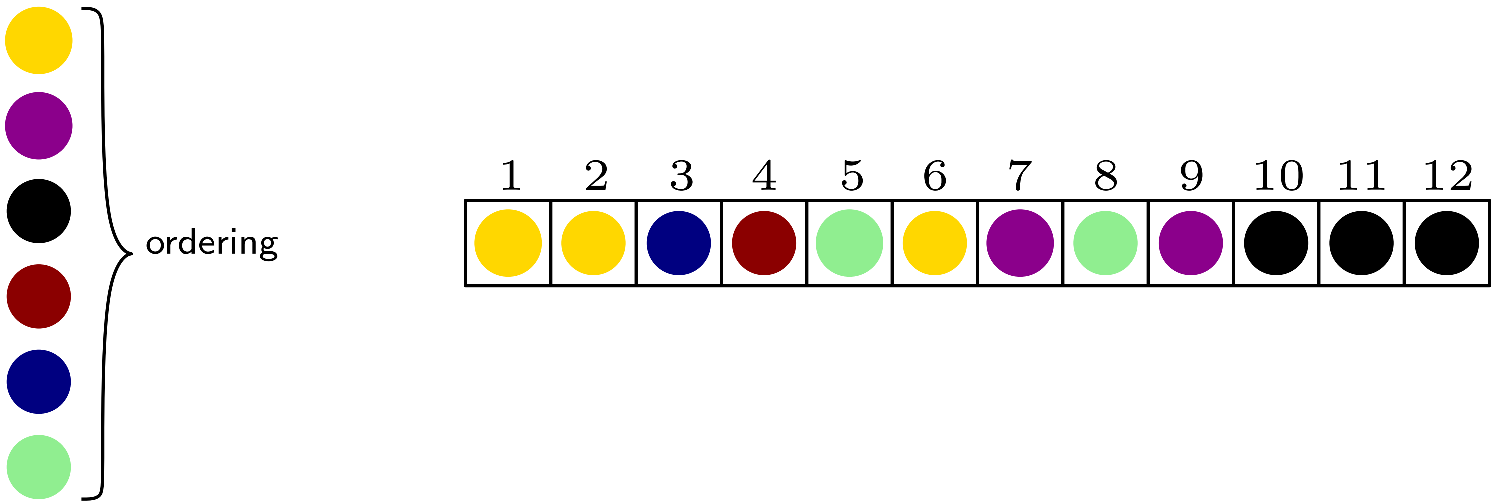
Algorithm: Preorder traversal of the tree

2 queries at node u:

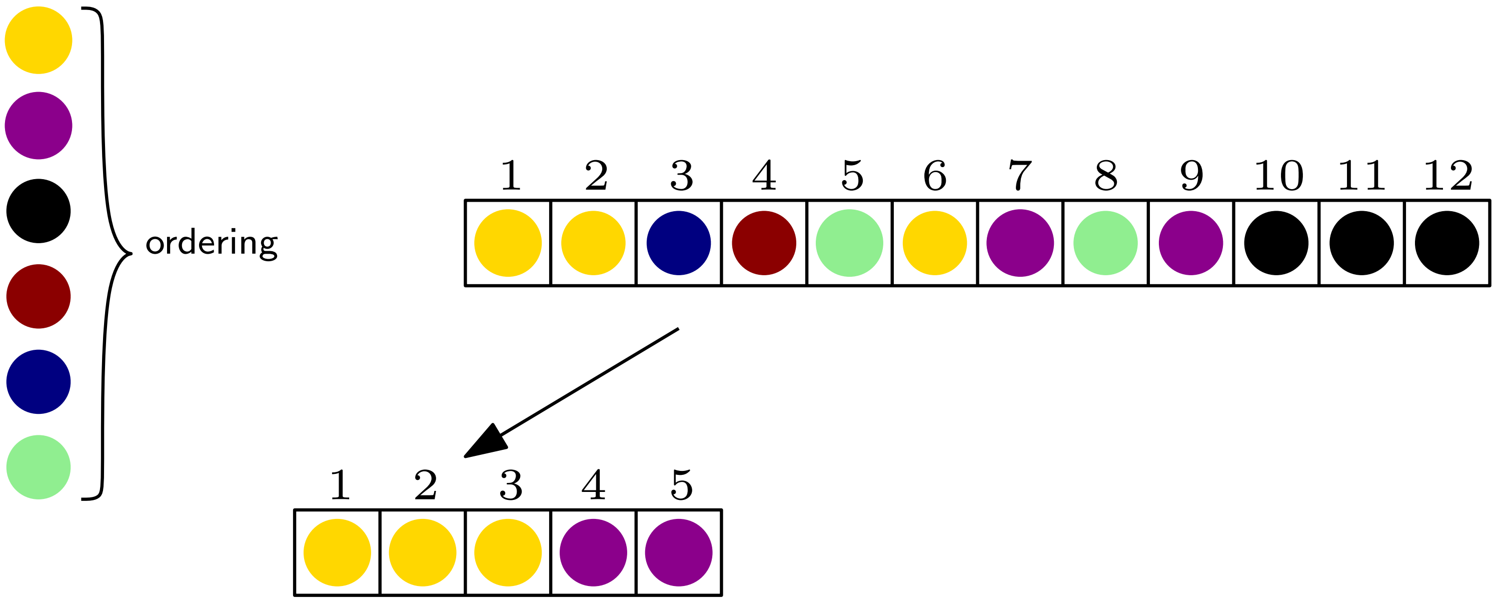
1. is color u in the output?
2. Is there a color in its subtree in the output?
    - Yes: Continue traversal
    - No: Skip subtree
    - Yes+output: Report result and skip subtree

Common Colors - Color Tree

Common Colors - Color Tree



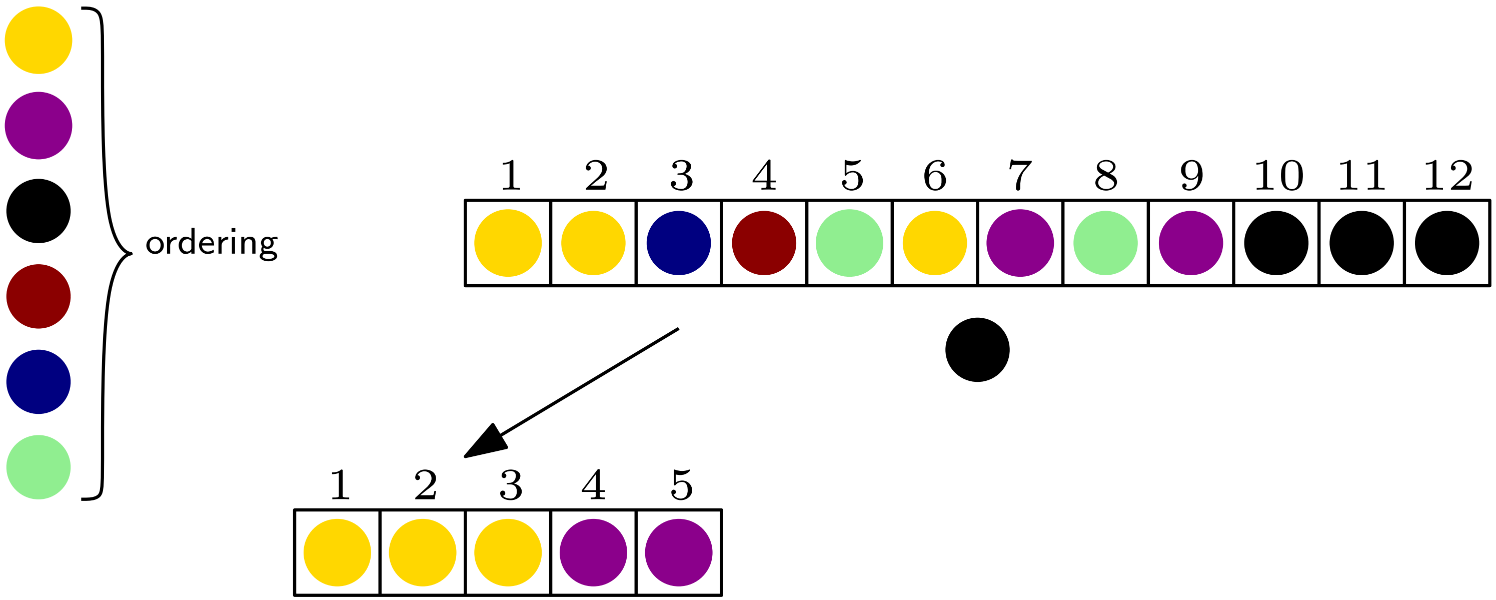
Common Colors - Color Tree



n_u := length of array at node u

Invariant: $n_{u_l}, n_{u_r} \leq \frac{1}{2} n_u$

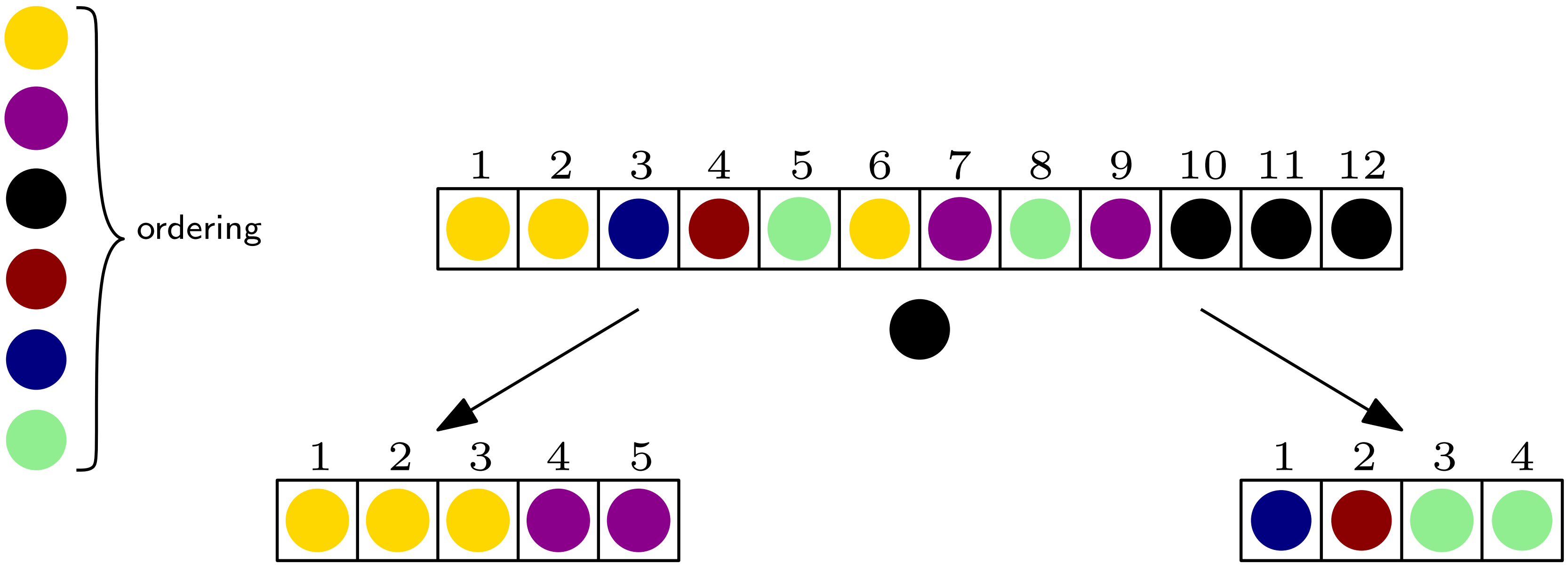
Common Colors - Color Tree



n_u := length of array at node u

Invariant: $n_{u_l}, n_{u_r} \leq \frac{1}{2} n_u$

Common Colors - Color Tree



n_u := length of array at node u

Invariant: $n_{u_l}, n_{u_r} \leq \frac{1}{2} n_u$

Common Colors - Query 2

Common Colors - Query 2

Is there a color from u's subtree in the output?

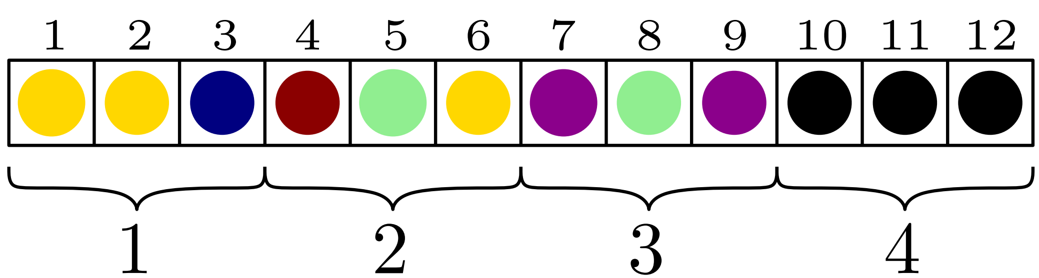
1. Yes: Continue traversal
2. No: Skip subtree
3. Yes+output: Report result and skip subtree

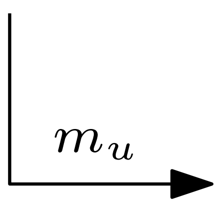
Common Colors - Query 2

Is there a color from u's subtree in the output?

1. Yes: Continue traversal
2. No: Skip subtree
3. Yes+output: Report result and skip subtree

Blocks:





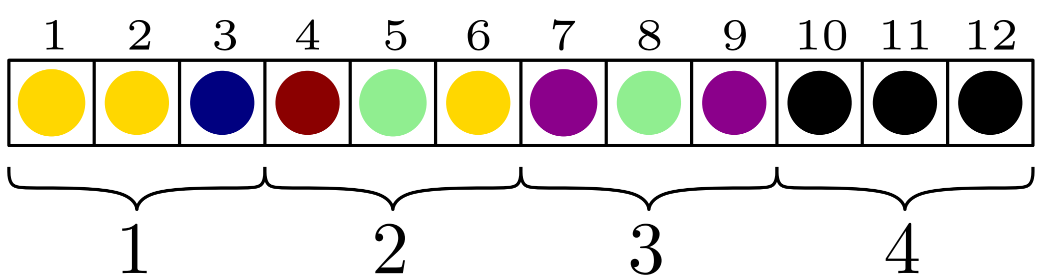
	1	2	3	4
1	1	1	0	0
2	1	1	1	0
3	0	1	1	0
4	0	0	0	1

(n_u entries in total)

Common Colors - Query 2

Is there a color from u's subtree in the output?

1. Yes: Continue traversal ($O(\log n / \log\log n)$)
2. No: Skip subtree
3. Yes+output: Report result and skip subtree



Blocks:

m_u

	1	2	3	4
1	1	1	0	0
2	1	1	1	0
3	0	1	1	0
4	0	0	0	1

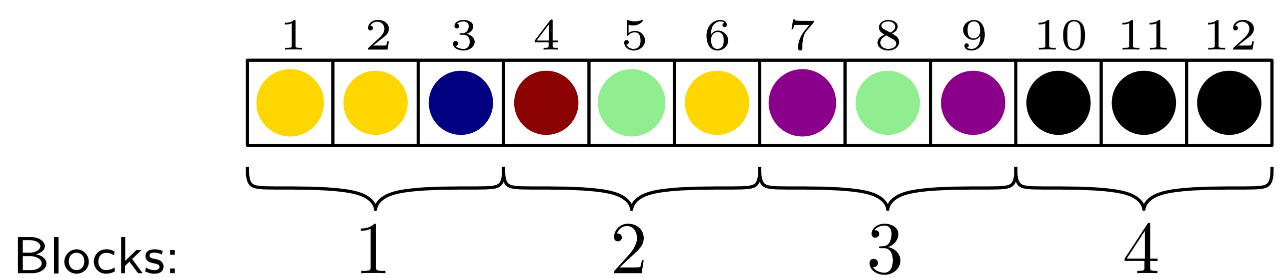
(n_u entries in total)

Common Colors - Query 2

Is there a color from u's subtree in the output?

1. Yes: Continue traversal ($O(\log n / \log\log n)$)
2. No: Skip subtree
3. Yes+output: Report result and skip subtree

$(O(\sqrt{n_u} \log^{\varepsilon} n))$ (applies to items 2 and 3)



m_u

	1	2	3	4
1	1	1	0	0
2	1	1	1	0
3	0	1	1	0
4	0	0	0	1

(n_u entries in total)

Common Colors - Time Analysis

Common Colors - Time Analysis

How much of the Color Tree is traversed?

Common Colors - Time Analysis

How much of the Color Tree is traversed?

Every ancestor of a leaf must have a color from its subtree in the ouput

Common Colors - Time Analysis

How much of the Color Tree is traversed?

Every ancestor of a leaf must have a color from its subtree in the ouput

Common Colors - Time Analysis

How much of the Color Tree is traversed?

Every ancestor of a leaf must have a color from its subtree in the ouput

Tree is of height $O(\log n)$

Common Colors - Time Analysis

How much of the Color Tree is traversed?

Every ancestor of a leaf must have a color from its subtree in the ouput

Tree is of height $O(\log n) \Rightarrow O(k \log n)$ nodes $+$ leaves are traversed

Common Colors - Time Analysis

How much of the Color Tree is traversed?

Every ancestor of a leaf must have a color from its subtree in the ouput

Tree is of height $O(\log n) \Rightarrow O(k \log n)$ nodes $+$ leaves are traversed

Time per internal node:

Common Colors - Time Analysis

How much of the Color Tree is traversed?

Every ancestor of a leaf must have a color from its subtree in the ouput

Tree is of height $O(\log n) \Rightarrow O(k \log n)$ nodes + leaves are traversed

Time per internal node: $O(\log n)$

Common Colors - Time Analysis

How much of the Color Tree is traversed?

Every ancestor of a leaf must have a color from its subtree in the ouput

Tree is of height $O(\log n) \Rightarrow O(k \log n)$ nodes $+$ leaves are traversed

Time per internal node: $O(\log n)$

Time per leaf u:

Common Colors - Time Analysis

How much of the Color Tree is traversed?

Every ancestor of a leaf must have a color from its subtree in the ouput

Tree is of height $O(\log n) \Rightarrow O(k \log n)$ nodes + leaves are traversed

Time per internal node: $O(\log n)$

Time per leaf u: $O(\sqrt{n_u} \log^{\varepsilon} n) = O\left(\sqrt{n \cdot (\frac{1}{2})^{\mathsf{depth}(u)}} \log^{\varepsilon} n\right)$

Common Colors - Time Analysis

How much of the Color Tree is traversed?

Every ancestor of a leaf must have a color from its subtree in the ouput

Tree is of height $O(\log n) \Rightarrow O(k \log n)$ nodes + leaves are traversed

Time per internal node: $O(\log n)$

Time per leaf u: $O(\sqrt{n_u} \log^{\varepsilon} n) = O\left(\sqrt{n \cdot (\frac{1}{2})^{\mathsf{depth}(u)}} \log^{\varepsilon} n\right)$

In total: $O\left(k \log^2 n + \sum_{\ell \in \mathsf{leaves}} \sqrt{n \cdot (\frac{1}{2})^{\mathsf{depth}(\ell)}} \log^{\varepsilon} n\right) =$

Common Colors - Time Analysis

How much of the Color Tree is traversed?

Every ancestor of a leaf must have a color from its subtree in the ouput

Tree is of height $O(\log n) \Rightarrow O(k \log n)$ nodes + leaves are traversed

Time per internal node: $O(\log n)$

Time per leaf u: $O(\sqrt{n_u} \log^{\varepsilon} n) = O\left(\sqrt{n \cdot (\frac{1}{2})^{\text{depth}(u)}} \log^{\varepsilon} n\right)$

In total: $O\left(k \log^2 n + \sum_{\ell \in \text{leaves}} \sqrt{n \cdot (\frac{1}{2})^{\text{depth}(\ell)}} \log^{\varepsilon} n\right) =$

$O\left(k \log^2 n + \sqrt{n} \log^{\varepsilon} n \sum_{\ell \in \text{leaves}} (\frac{1}{2})^{\text{depth}(\ell)/2}\right) =$

Common Colors - Time Analysis

How much of the Color Tree is traversed?

Every ancestor of a leaf must have a color from its subtree in the ouput

Tree is of height $O(\log n) \Rightarrow O(k \log n)$ nodes + leaves are traversed

Time per internal node: $O(\log n)$

Time per leaf u: $O(\sqrt{n_u} \log^{\varepsilon} n) = O\left(\sqrt{n \cdot (\frac{1}{2})^{\text{depth}(u)}} \log^{\varepsilon} n\right)$

In total: $O\left(k \log^2 n + \sum_{\ell \in \text{leaves}} \sqrt{n \cdot (\frac{1}{2})^{\text{depth}(\ell)}} \log^{\varepsilon} n\right) =$

$O\left(k \log^2 n + \sqrt{n} \log^{\varepsilon} n \sum_{\ell \in \text{leaves}} (\frac{1}{2})^{\text{depth}(\ell)/2}\right) =$

$O\left(k \log^2 n + \sqrt{n} \log^{\varepsilon} n \sum_{\ell \in \text{leaves}} \sqrt{1} (\frac{1}{2})^{\text{depth}(\ell)/2}\right) =$

Common Colors - Time Analysis

How much of the Color Tree is traversed?

Every ancestor of a leaf must have a color from its subtree in the ouput

Tree is of height $O(\log n) \Rightarrow O(k \log n)$ nodes + leaves are traversed

Time per internal node: $O(\log n)$

Time per leaf u: $O(\sqrt{n_u} \log^{\varepsilon} n) = O\left(\sqrt{n \cdot (\frac{1}{2})^{\text{depth}(u)}} \log^{\varepsilon} n\right)$

In total: $O\left(k \log^2 n + \sum_{\ell \in \text{leaves}} \sqrt{n \cdot (\frac{1}{2})^{\text{depth}(\ell)}} \log^{\varepsilon} n\right) =$

$O\left(k \log^2 n + \sqrt{n} \log^{\varepsilon} n \sum_{\ell \in \text{leaves}} (\frac{1}{2})^{\text{depth}(\ell)/2}\right) =$

$O\left(k \log^2 n + \sqrt{n} \log^{\varepsilon} n \sum_{\ell \in \text{leaves}} \sqrt{1} (\frac{1}{2})^{\text{depth}(\ell)/2}\right) =$

$O\left(k \log^2 n + \sqrt{n} \log^{\varepsilon} n \sqrt{\sum_{\ell \in \text{leaves}} 1 \sum_{\ell \in \text{leaves}} (\frac{1}{2})^{\text{depth}(\ell)}}\right) =$

Common Colors - Time Analysis

How much of the Color Tree is traversed?

Every ancestor of a leaf must have a color from its subtree in the ouput

Tree is of height $O(\log n) \Rightarrow O(k \log n)$ nodes + leaves are traversed

Time per internal node: $O(\log n)$

Time per leaf u: $O(\sqrt{n_u} \log^{\varepsilon} n) = O\left(\sqrt{n \cdot (\frac{1}{2})^{\text{depth}(u)}} \log^{\varepsilon} n\right)$

In total: $O\left(k \log^2 n + \sum_{\ell \in \text{leaves}} \sqrt{n \cdot (\frac{1}{2})^{\text{depth}(\ell)}} \log^{\varepsilon} n\right) =$

$O\left(k \log^2 n + \sqrt{n} \log^{\varepsilon} n \sum_{\ell \in \text{leaves}} (\frac{1}{2})^{\text{depth}(\ell)/2}\right) =$

$O\left(k \log^2 n + \sqrt{n} \log^{\varepsilon} n \sum_{\ell \in \text{leaves}} \sqrt{1} (\frac{1}{2})^{\text{depth}(\ell)/2}\right) =$

$O\left(k \log^2 n + \sqrt{n} \log^{\varepsilon} n \sqrt{\sum_{\ell \in \text{leaves}} 1 \sum_{\ell \in \text{leaves}} (\frac{1}{2})^{\text{depth}(\ell)}}\right) =$

$O\left(k \log^2 n + \sqrt{n} \log^{\varepsilon} n \sqrt{k \log n}\right) = O\left(k \log^2 n + \sqrt{nk} \log^{1/2+\varepsilon} n\right)$

Common Colors - Space Analysis

Common Colors - Space Analysis

Color Tree: $\leq n$ nodes and $\leq \log n$ levels

Common Colors - Space Analysis

Color Tree: $\leq n$ nodes and $\leq \log n$ levels

Each level uses $O(n)$ bits for the rank/select structures

Common Colors - Space Analysis

Color Tree: $\leq n$ nodes and $\leq \log n$ levels

Each level uses $O(n)$ bits for the rank/select structures

$\Rightarrow O(n \log n)$ bits $= O(n)$ words

Common Colors - Space Analysis

Color Tree: $\leq n$ nodes and $\leq \log n$ levels

Each level uses $O(n)$ bits for the rank/select structures

$\Rightarrow O(n \log n)$ bits $= O(n)$ words

Each level uses $O(n)$ bits for their matrices and RMQ

Common Colors - Space Analysis

Color Tree: $\leq n$ nodes and $\leq \log n$ levels

Each level uses $O(n)$ bits for the rank/select structures
$\Rightarrow O(n \log n)$ bits $= O(n)$ words

Each level uses $O(n)$ bits for their matrices and RMQ
$\Rightarrow O(n \log n)$ bits $= O(n)$ words

Common Colors - Space Analysis

Color Tree: $\leq n$ nodes and $\leq \log n$ levels

Each level uses $O(n)$ bits for the rank/select structures

$\Rightarrow O(n \log n)$ bits $= O(n)$ words

Each level uses $O(n)$ bits for their matrices and RMQ

$\Rightarrow O(n \log n)$ bits $= O(n)$ words

Range emptiness for each color: $O(n)$ words in total

Common Colors - Space Analysis

Color Tree: $\leq n$ nodes and $\leq \log n$ levels

Each level uses $O(n)$ bits for the rank/select structures
$\Rightarrow O(n \log n)$ bits $= O(n)$ words
Each level uses $O(n)$ bits for their matrices and RMQ
$\Rightarrow O(n \log n)$ bits $= O(n)$ words

Range emptiness for each color: $O(n)$ words in total

Range predecessor/successor: $O(n)$ words in total

Common Colors - Space Analysis

Color Tree: $\leq n$ nodes and $\leq \log n$ levels

Each level uses $O(n)$ bits for the rank/select structures
$\Rightarrow O(n \log n)$ bits $= O(n)$ words

Each level uses $O(n)$ bits for their matrices and RMQ
$\Rightarrow O(n \log n)$ bits $= O(n)$ words

Range emptiness for each color: $O(n)$ words in total

Range predecessor/successor: $O(n)$ words in total

The structure uses linear space, i.e. $O(n)$ words

Thank You