9 Range-Minimum Queries

25 April 2022

Sebastian Wild

COMP526 (Spring 2022)
University of Liverpool
version `2022-04-24 20:13`

Learning Outcomes

1. Know the *RMQ problem* and its *connection* to longest common extensions in strings.
2. Know and understand trivial RMQ solutions and *sparse tables*.
3. Know and understand the *Cartesian trees* data structure.
4. Know and understand the *exhaustive-tabulation technique* for RMQ with linear-time preprocessing.

Unit 9: *Range-Minimum Queries*

Outline

9 Range-Minimum Queries

9.1 Introduction

Range-minimum queries (RMQ)

array / numbers don't change

- **Given:** Static array $A[0..n)$ of numbers
- **Goal:** Find minimum in a range;
  A known in advance and can be preprocessed

$\mathrm{RMQ}(7, 15) = 10$

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	4	6	4	7	10	5	6	3	11	2	2	3	6	10	9	13	4	6	16	10

- **Nitpicks:**
  - Report *index* of minimum, not its value
  - Report *leftmost* position in case of ties

Clicker Question

Given the array from the slides, what is $\mathrm{RMQ}_A(1,6)$

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	4	6	4	7	10	5	6	3	11	2	2	3	6	10	9	13	4	6	16	10

sli.do/comp526

Rules of the Game

- comparison-based $\rightsquigarrow$ values don't matter, only relative order

- Two main quantities of interest:

  $\rightsquigarrow$ space usage $\leq P(n)$

  1. **Preprocessing time**: Running time $P(n)$ of the preprocessing step
  2. **Query time**: Running time $Q(n)$ of one query (using precomputed data)

- Write $\langle P(n), Q(n) \rangle$ **time solution** for short

Clicker Question

What do you think, what running times can we achieve? For a $\langle P(n), Q(n)\rangle$ time solution, enter “`<P(n),Q(n)>`”.

`sli.do/comp526`

9.2 RMQ, LCP, LCE, LCA — WTF?

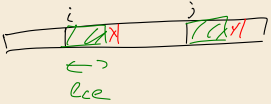
Recall Unit 6

Application 4: Longest Common Extensions

- We implicitly used a special case of a more general, versatile idea:

  The ***longest common extension (LCE)*** data structure:

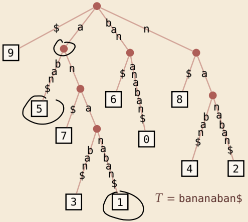
  - **Given:** String $T[0..n-1]$
  - **Goal:** Answer LCE queries, i. e.,
    given positions i, j in T,
    how far can we read the same text from there?
    formally: $\mathrm{LCE}(i, j) = \max\{\ell : T[i..i+\ell) = T[j..j+\ell)\}$



$\rightsquigarrow$ use suffix tree of T!

longest common prefix of ith and jth suffix

- In $\mathcal{T}$: $\mathrm{LCE}(i, j) = \mathrm{LCP}(T_i, T_j)$ $\rightsquigarrow$ same thing, different name!
  $=$ string depth of
  lowest common ancester (LCA) of
  leaves $\boxed{i}$ and $\boxed{j}$



- in short: $\boxed{\mathrm{LCE}(i, j) = \mathrm{LCP}(T_i, T_j) = \mathrm{stringDepth}\big(\mathrm{LCA}(\boxed{i}, \boxed{j})\big)}$

Recall Unit 6

Efficient LCA

How to find lowest common ancestors?

- Could walk up the tree to find LCA $\rightsquigarrow$ $\Theta(n)$ worst case
- Could store all LCAs in big table $\rightsquigarrow$ $\Theta(n^2)$ space and preprocessing

Amazing result: Can compute data structure in $\Theta(n)$ time and space that finds any LCA is **constant(!) time**.

- a bit tricky to understand
- but a theoretical breakthrough
- and useful in practice

and suffix tree construction inside . . .

$\rightsquigarrow$ for now, use $O(1)$ LCA as black box.

$\rightsquigarrow$ After linear preprocessing (time & space), we can find LCEs in $O(1)$ time.

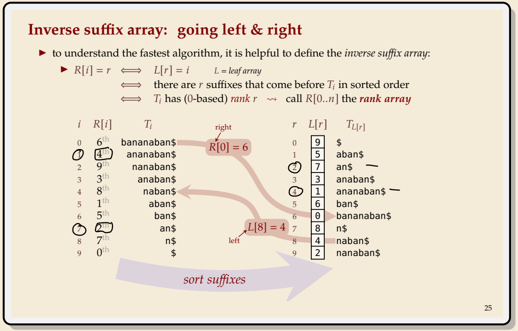
16

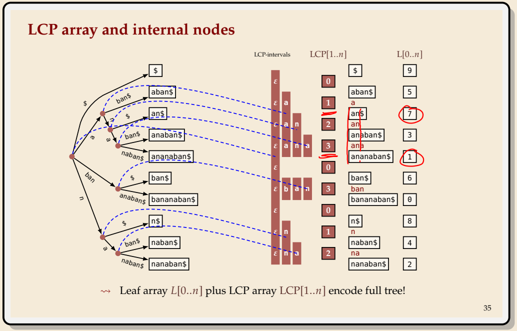
Finally: Longest common extensions

- In Unit 6: Left question open how to compute LCA in suffix trees
- But: Enhanced Suffix Array makes life easier!

0 1 2 3 4 5 6 7 8 9
b a n a n a b a n $

$$\mathrm{LCE}(\overset{1}{i}, \overset{7}{j}) \;=\; \mathrm{LCP}\big[\mathrm{RMQ}_{\mathrm{LCP}}(\min\{R[i], R[j]\} + 1, \max\{R[i], R[j]\})\big]$$





RMQ Implications for LCE

- Recall: Can compute (inverse) suffix array and LCP array in $O(n)$ time

$\rightsquigarrow$ A $\langle P(n), Q(n)\rangle$ time RMQ data structure implies a $\langle P(n), Q(n)\rangle$ time solution for longest-common extensions

$+ O(n)$

9.3 Trivial Solutions & Sparse Tables

Trivial Solutions

$\mathrm{RMQ}(7, 15) = 10$

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	4	6	4	7	10	5	6	3	11	2	2	3	6	10	9	13	4	6	16	10

- Two easy solutions show extreme ends of scale:

Trivial Solutions

$\mathrm{RMQ}(7, 15) = 10$

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	4	6	4	7	10	5	6	3	11	2	2	3	6	10	9	13	4	6	16	10

- Two easy solutions show extreme ends of scale:

1. **Scan on demand**

  - no preprocessing at all
  - answer $\mathrm{RMQ}(i, j)$ by scanning through $A[i..j]$, keeping track of min

  $\rightsquigarrow$ $\langle O(1), O(n) \rangle$

Trivial Solutions

RMQ(7, 15) = 10

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	4	6	4	7	10	5	6	3	11	2	2	3	6	10	9	13	4	6	16	10

- Two easy solutions show extreme ends of scale:

1. **Scan on demand**
   - no preprocessing at all
   - answer $\mathrm{RMQ}(i, j)$ by scanning through $A[i..j]$, keeping track of min

   $\rightsquigarrow$ $\langle O(1), O(n)\rangle$

2. **Precompute all**
   - Precompute all answers in a big 2D array $M[0..n)[0..n)$
   - queries simple: $\mathrm{RMQ}(i, j) = M[i][j]$

   $\rightsquigarrow$ $\langle O(n^3), O(1)\rangle$ —— fill $\Theta(n^2)$ cells, each takes $O(n)$

Trivial Solutions

$\mathrm{RMQ}(7, 15) = 10$

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	4	6	4	7	10	5	6	3	11	2	2	3	6	10	9	13	4	6	16	10

i $\qquad\qquad$ j

i $\qquad\qquad$ $j-1$

- Two easy solutions show extreme ends of scale:

1. **Scan on demand**
   - no preprocessing at all
   - answer $\mathrm{RMQ}(i, j)$ by scanning through $A[i..j]$, keeping track of min

   $\rightsquigarrow \langle O(1), O(n) \rangle$

2. **Precompute all**
   - Precompute all answers in a big 2D array $M[0..n)[0..n)$
   - queries simple: $\mathrm{RMQ}(i, j) = M[i][j]$

   $\rightsquigarrow \langle O(n^3), O(1) \rangle$
   - Preprocessing can reuse partial results $\rightsquigarrow \langle O(n^2), O(1) \rangle$

$M[i][j] = \arg\min\{A[M[i][j-1]], A[j]\}$

if $A[M[i][j-1]] \le A[j]$

then $M[i][j-1]$

else j

Sparse Table

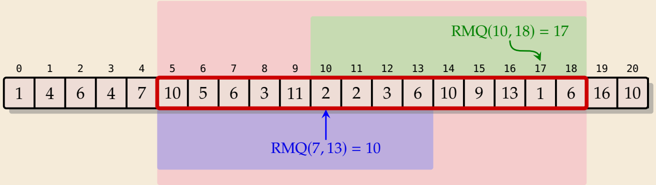
- **Idea:** Like "precompute-all", but keep only some entries
- store $M[i][j]$ iff $\ell = j - i + 1$ is 2^k.
  - $\rightsquigarrow$ $\leq n \cdot \lg n$ entries
  - $\rightsquigarrow$ Can be stored as $M'[i][k]$ $= M[i][i + 2^k - 1]$

Sparse Table

- **Idea:** Like "precompute-all", but keep only some entries
- store $M[i][j]$ iff $\ell = j - i + 1$ is 2^k.
  - $\rightsquigarrow$ $\le n \cdot \lg n$ entries
  - $\rightsquigarrow$ Can be stored as $M'[i][k]$
- How to answer queries?

Sparse Table

- **Idea:** Like "precompute-all", but keep only some entries
- store $M[i][j]$ iff $\ell = j - i + 1$ is 2^k.
  - $\leadsto \le n \cdot \lg n$ entries
  - $\leadsto$ Can be stored as $M'[i][k]$
- How to answer queries?



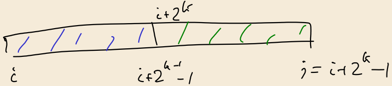
1. Find k with $\ell/2 \le 2^k \le \ell$
2. Cover range $[i..j]$ by
   2^k positions right from i and
   2^k positions left from j
3. $\mathrm{RMQ}(i, j) =$
   $\arg\min\{A[rmq_1], A[rmq_2]\}$

   with $rmq_1 = \mathrm{RMQ}(i, i + 2^k - 1)$
   $rmq_2 = \mathrm{RMQ}(j - 2^k + 1, j)$

$rmq_1 = M'[i][k]$

$rmq_2 = M'[j - 2^k + 1][k]$

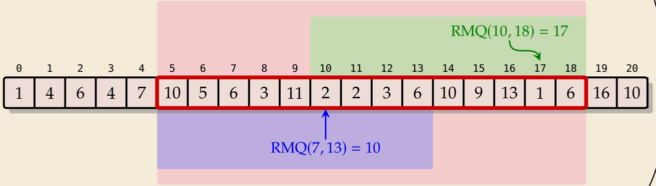
Sparse Table



- **Idea:** Like "precompute-all", but keep only some entries
- store $M[i][j]$ iff $\ell = j - i + 1$ is 2^k.
  - $\rightsquigarrow$ $\leq n \cdot \lg n$ entries
  - $\rightsquigarrow$ Can be stored as $M'[i][k]$

$M'[i][k] =$
$\arg\min \{ A[\underline{M[i][k-1]}], A[\underline{M[i+2^k][k-1]}] \}$

- How to answer queries?



1. Find k with $\ell/2 \leq 2^k \leq \ell$
2. Cover range $[i..j]$ by
   2^k positions right from i and
   2^k positions left from j
3. $\text{RMQ}(i, j) =$
   $\arg\min\{A[rmq_1], A[rmq_2]\}$

   with $rmq_1 = \text{RMQ}(i, i + 2^k - 1)$
   $rmq_2 = \text{RMQ}(j - 2^k + 1, j)$

- Preprocessing can be done in $O(n \log n)$ times

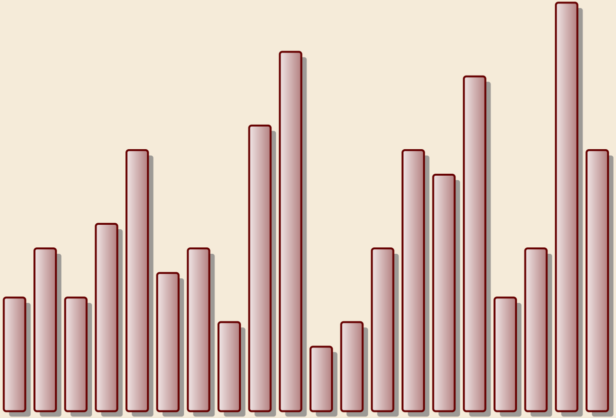
$\rightsquigarrow$ $\langle O(n \log n), O(1) \rangle$ time solution!

9.4 Cartesian Trees

RMQ & LCA

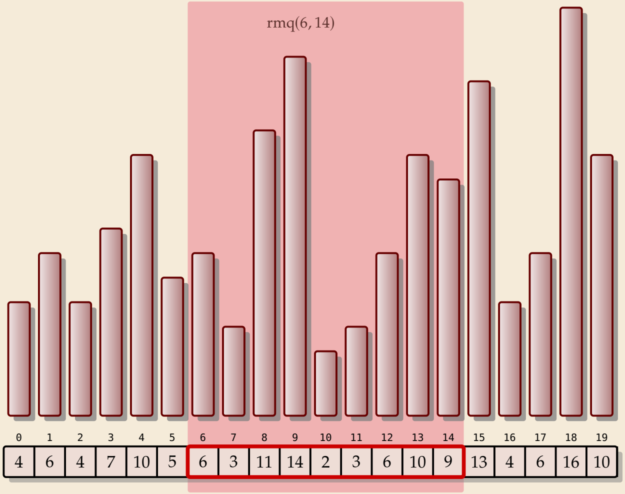
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
4	6	4	7	10	5	6	3	11	14	2	3	6	10	9	13	4	6	16	10

RMQ & LCA



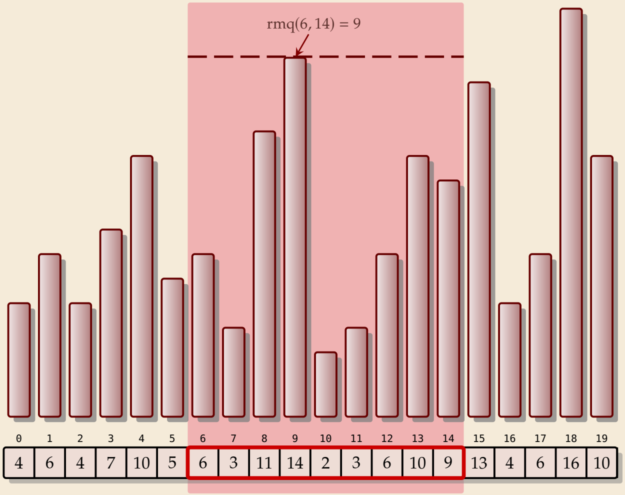
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
4	6	4	7	10	5	6	3	11	14	2	3	6	10	9	13	4	6	16	10

RMQ & LCA



- **Range-max queries** on array A:
  $\mathrm{rmq}_A(i, j) = \underset{i \le k \le j}{\arg\max}\, A[k]$
  $= \textit{index}$ of max

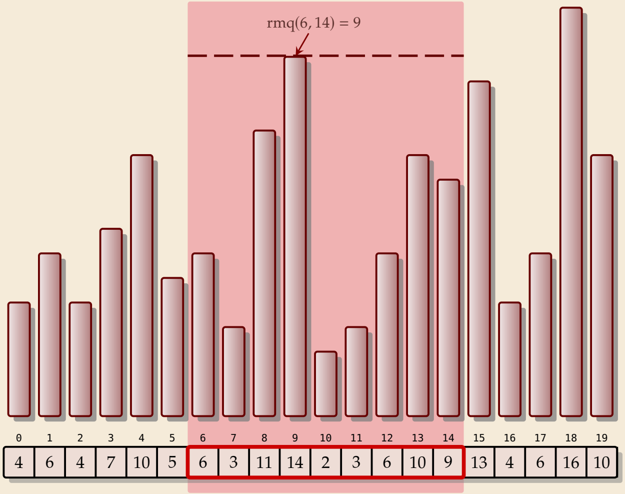
RMQ & LCA



- **R**ange-**m**ax **q**ueries on array A:

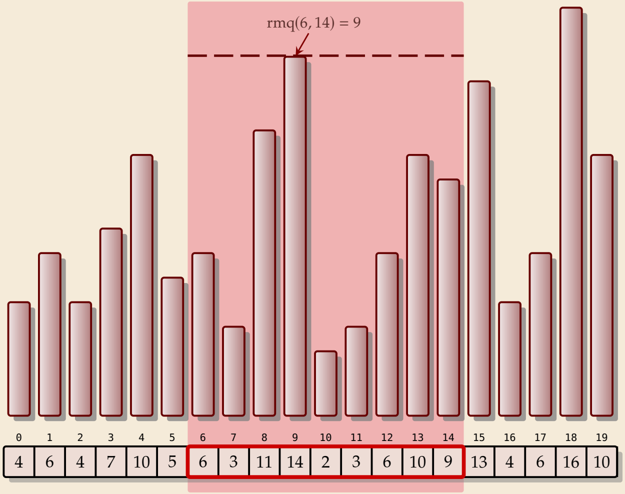
$$\mathrm{rmq}_A(i, j) = \arg\max_{i \le k \le j} A[k]$$
$$= \textit{index} \text{ of max}$$

RMQ & LCA



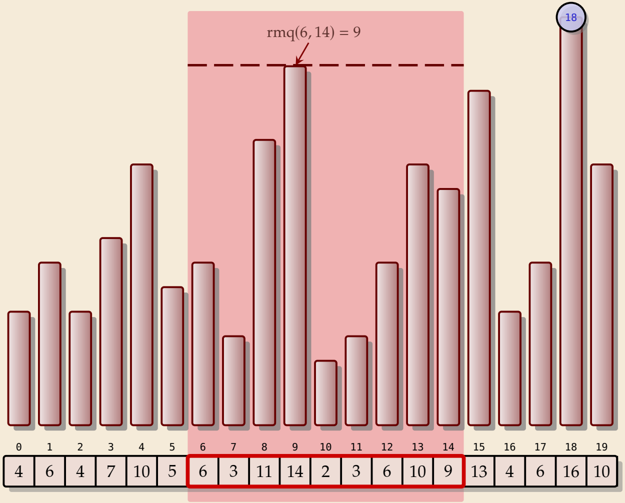
- **Range-max queries** on array A:
  $$\mathrm{rmq}_A(i,j) = \underset{i \le k \le j}{\arg\max}\, A[k]$$
  $= \textit{index}$ of max
- **Task:** Preprocess A,
  then answer RMQs fast

RMQ & LCA



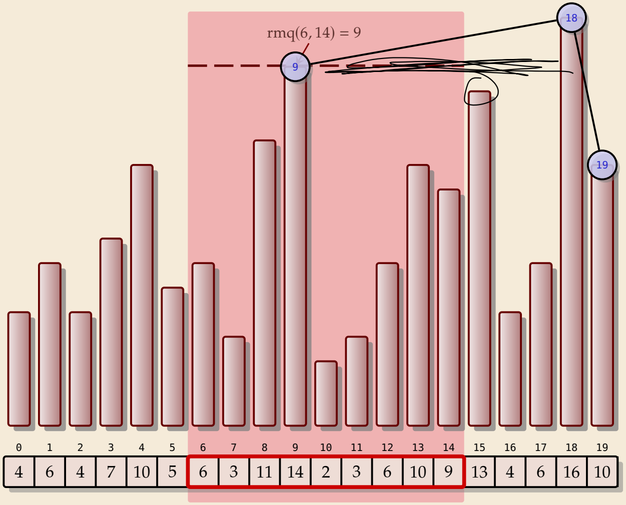
- **Range-max queries** on array A:
  $$\operatorname{rmq}_A(i, j) = \underset{i \le k \le j}{\arg\max}\, A[k]$$
  $$= \textit{index} \text{ of max}$$
- **Task:** Preprocess A, then answer RMQs fast ideally constant time!

RMQ & LCA



- **Range-max queries** on array A:
  $$\text{rmq}_A(i, j) = \arg\max_{i \le k \le j} A[k]$$
  $= \textit{index}$ of max
- **Task:** Preprocess A, then answer RMQs fast ideally constant time!
- **Cartesian tree:** (cf. *treap*) construct binary tree by sweeping line down

RMQ & LCA

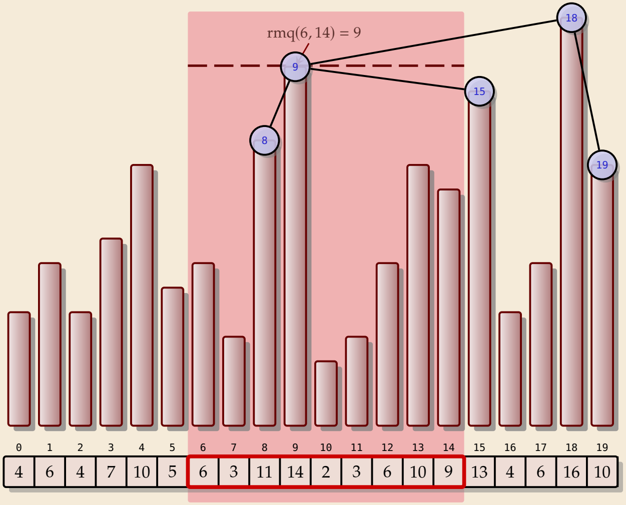


- **Range-max queries** on array A:
  $\text{rmq}_A(i, j) = \underset{i \le k \le j}{\arg\max} A[k]$
  $= \textit{index}$ of max

- **Task:** Preprocess A,
  then answer RMQs fast
  ideally constant time!

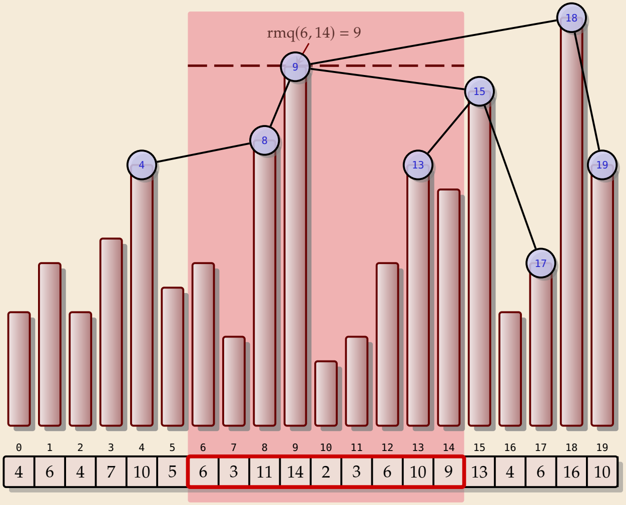
- **Cartesian tree:** (cf. *treap*)
  construct binary tree by
  sweeping line down

RMQ & LCA



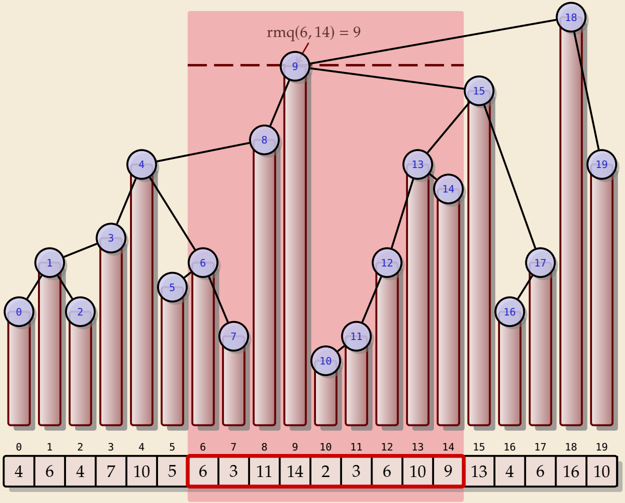
- **Range-max queries** on array A:
  $\mathrm{rmq}_A(i, j) = \arg\max_{i \le k \le j} A[k]$
  $= \textit{index}$ of max
- **Task:** Preprocess A,
  then answer RMQs fast
  ideally constant time!
- **Cartesian tree:** (cf. *treap*)
  construct binary tree by
  sweeping line down

RMQ & LCA



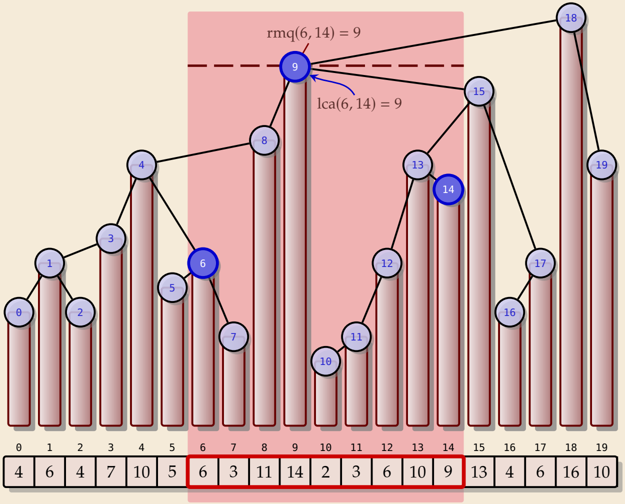
- **Range-max queries** on array A:
  $\text{rmq}_A(i, j) = \arg\max_{i \le k \le j} A[k]$
  $= \textit{index}$ of max
- **Task:** Preprocess A,
  then answer RMQs fast
  ideally constant time!
- **Cartesian tree:** (cf. *treap*)
  construct binary tree by
  sweeping line down

RMQ & LCA



- **Range-max queries** on array A:
  $\mathrm{rmq}_A(i,j) = \underset{i \le k \le j}{\arg\max}\, A[k]$
  $= \textit{index}$ of max
- **Task:** Preprocess A,
  then answer RMQs fast
  ideally constant time!
- **Cartesian tree:** (cf. *treap*)
  construct binary tree by
  sweeping line down

RMQ & LCA



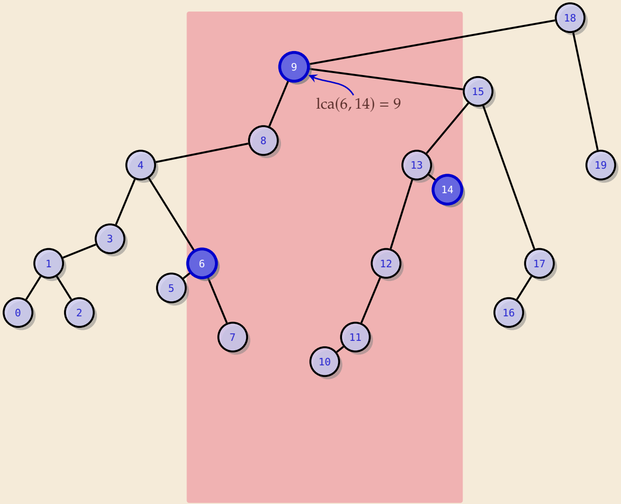
- **Range-max queries** on array A:
  $\mathrm{rmq}_A(i, j) = \arg\max_{i \le k \le j} A[k]$
  $= \textit{index}$ of max

- **Task:** Preprocess A,
  then answer RMQs fast
  ideally constant time!

- **Cartesian tree:** (cf. *treap*)
  construct binary tree by
  sweeping line down

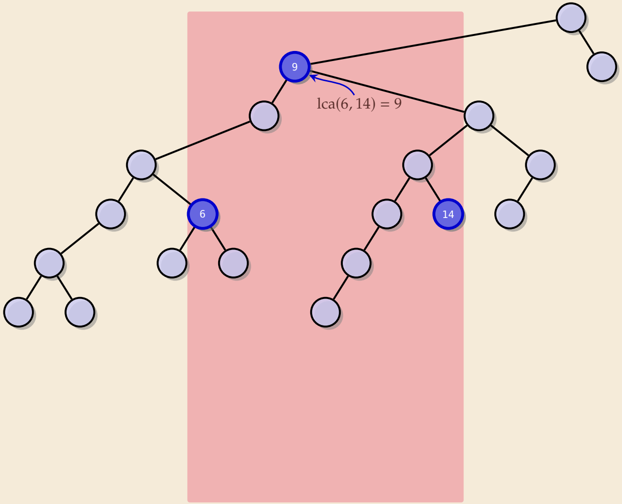
- $\mathrm{rmq}(i, j)$ =
  lowest common ancestor (LCA)

RMQ & LCA



- **Range-max queries** on array A:
  $\mathrm{rmq}_A(i,j) = \arg\max_{i \le k \le j} A[k]$
  $= \textit{index}$ of max
- **Task:** Preprocess A,
  then answer RMQs fast
  ideally constant time!
- **Cartesian tree:** (cf. *treap*)
  construct binary tree by
  sweeping line down
- $\mathrm{rmq}(i,j)$ =
  lowest common ancestor (LCA)

RMQ & LCA



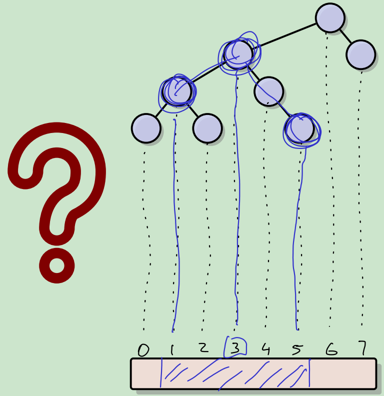
- **Range-max queries** on array A:
  $\mathrm{rmq}_A(i, j) = \underset{i \le k \le j}{\arg\max}\, A[k]$
  $= \textit{index}$ of max

- **Task:** Preprocess A,
  then answer RMQs fast
  ideally constant time!

- **Cartesian tree:** (cf. *treap*)
  construct binary tree by
  sweeping line down

- $\mathrm{rmq}(i, j)$ = inorder of
  lowest common ancestor (LCA)
  of ith and jth node in inorder

Clicker Question



Given the (max-oriented) Cartesian tree for A on the left, what is $\mathrm{RMQ}_A(1,5)$?

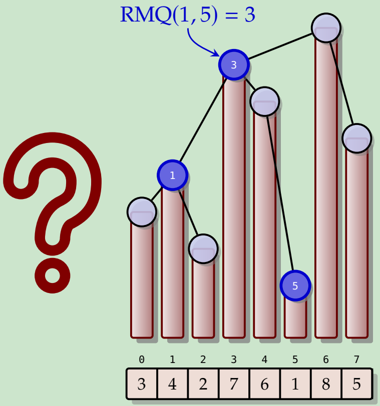
sli.do/comp526

Clicker Question

Given the (max-oriented) Cartesian tree for A on the left, what is $\mathrm{RMQ}_A(1,5)$?

sli.do/comp526

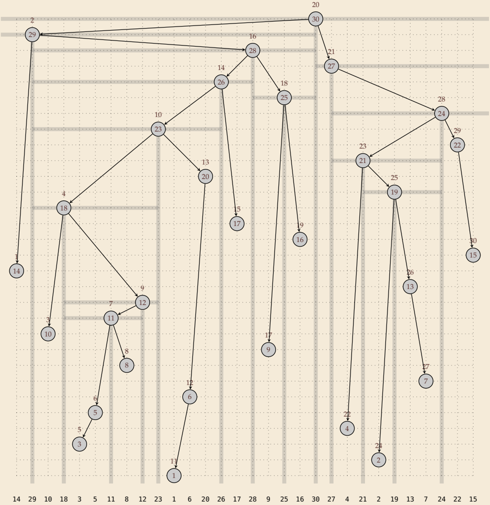
Clicker Question



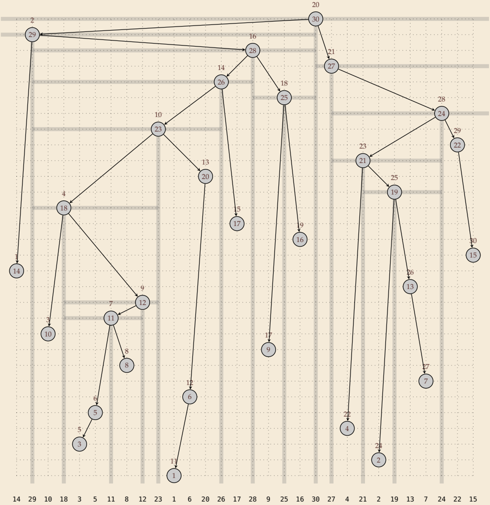
Given the (max-oriented) Cartesian tree for A on the left, what is $\mathrm{RMQ}_A(1,5)$?

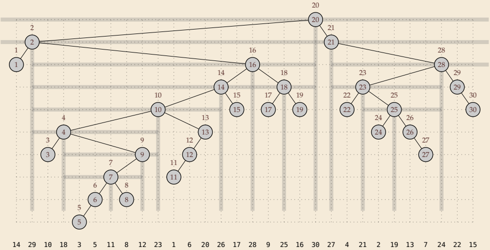
`sli.do/comp526`

Cartesian Tree – Larger Example

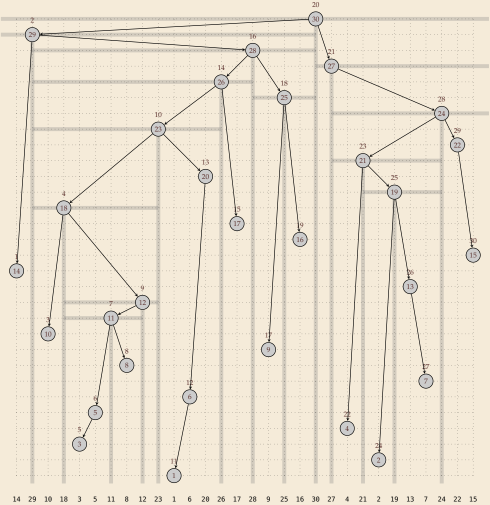


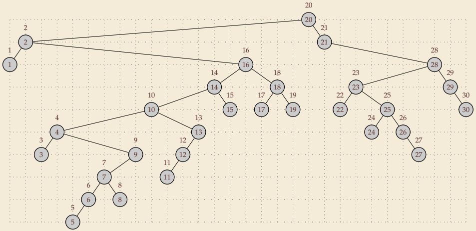
Cartesian Tree – Larger Example



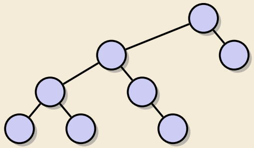


Cartesian Tree – Larger Example





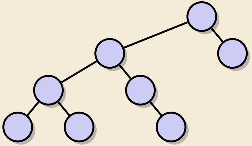
Counting binary trees



- Given the Cartesian tree,
  all RMQ answers are determined

  and vice versa!

Counting binary trees



- Given the Cartesian tree, all RMQ answers are determined

  and vice versa!

naive bound: n^{n^2}

$\lg(n^{n^2}) = n^2 \lg n$

- How many different Cartesian trees are there for arrays of length n?
  - known result: *Catalan numbers* $\frac{1}{n+1}\binom{2n}{n}$
  - easy to see: $\leq 2^{2n}$

visit all nodes in preorder

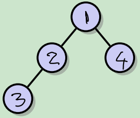
store for visited node:

(has left child, has right child)

$\rightsquigarrow$ many arrays will give rise to the same Cartesian tree

Can we exploit that?

Clicker Question

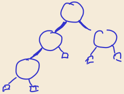


What binary string corresponds to the tree shown on the left?

(using the encoding just discussed)

1 2 3 4

11 10 00 00



visit all nodes in preorder

store for visited node:

(has left child, has right child)

sli.do/comp526