Applied Algorithms

CSCI-B505 / INFO-I500

Lecture 15.

Selection and Beyond

with

Rank/Select and Wavelet Tree Data Structures

M. Oguzhan Kulekci

- Predecessor/Successor with R/S
- Range Quantile Queries

Predecessor/Successor

Given a sequence of items,

- Predecessor: Find the largest item less than a queried one
- Successor :Find the smallest item larger than a queried one.

$$S = \{5,7,3,10,9,1,20,8\}$$

$$Pred(3) = 1 \quad Succ(3) = 5$$

$$Pred(10) = 9 \quad Succ(10) = 20$$

Predecessor/Successor

- Keeping a sorted copy of the array with some auxiliary data can solve it. Efficient ??
- There are many different data structures for such queries
- We will focus on a bit-vector solution, assuming we have R/S support

$$S = \{5,7,3,10,9,1,20,8\}$$

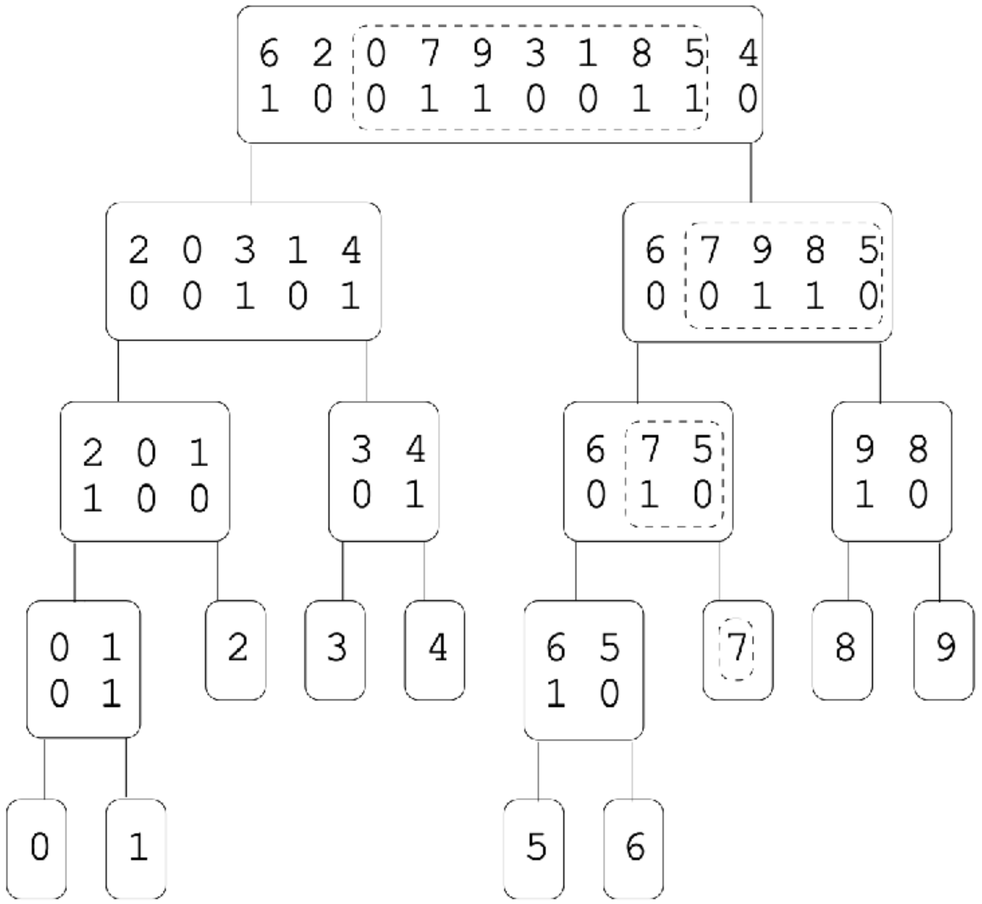
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	0	1	0	1	0	1	1	1	1	0	0	0	0	0	0	0	0	0	1

- If BV[k] is not 1, then **`Pred(k)=select(rank(k))`**,

  else **`Pred(k)=select(rank(k)-1)`**

- **`Succ(k)=select(rank(k)+1)`**

Range Quantile Queries with WT



$k = 5$
$\ell = 3$
$r = 9$

$k = 2$
$\ell = 2$
$r = 5$

$k = 2$
$\ell = 2$
$r = 3$

$k = 1$
$\ell = 1$
$r = 1$

Time Complexity: $O(\log n)$

Space complexity: $O(n \log n)$

Fig. 1. A wavelet tree T (left) for $s = 6, 2, 0, 7, 9, 3, 1, 8, 5, 4$, and the values (right) the variables k, ℓ and r take on as we search for the 5th smallest element in $s[3..9]$. The dashed boxes in T show the ranges from which we recursively select.

See https://arxiv.org/pdf/0903.4726.pdf

Compact Integer Codes with R/S Dictionary

$$X = \langle 3,6,11,5,1,3,15,9,13 \rangle$$

01100110000101101001011011000111100010010001101

L = **0100100010110100010010001**

P = 110011001111111001101

What is the third integer, given L and P ?

select(3,L) = 9

select(2,L) = 5

Bits 4,5,6 on P = 011

→ 0001011 = 11

1	**1**
3	0**11**
5	00**1**01
6	00**11**0
9	000**1**001
11	000**1**011
13	000**1**101
15	000**1**111

Elias Integer Codes

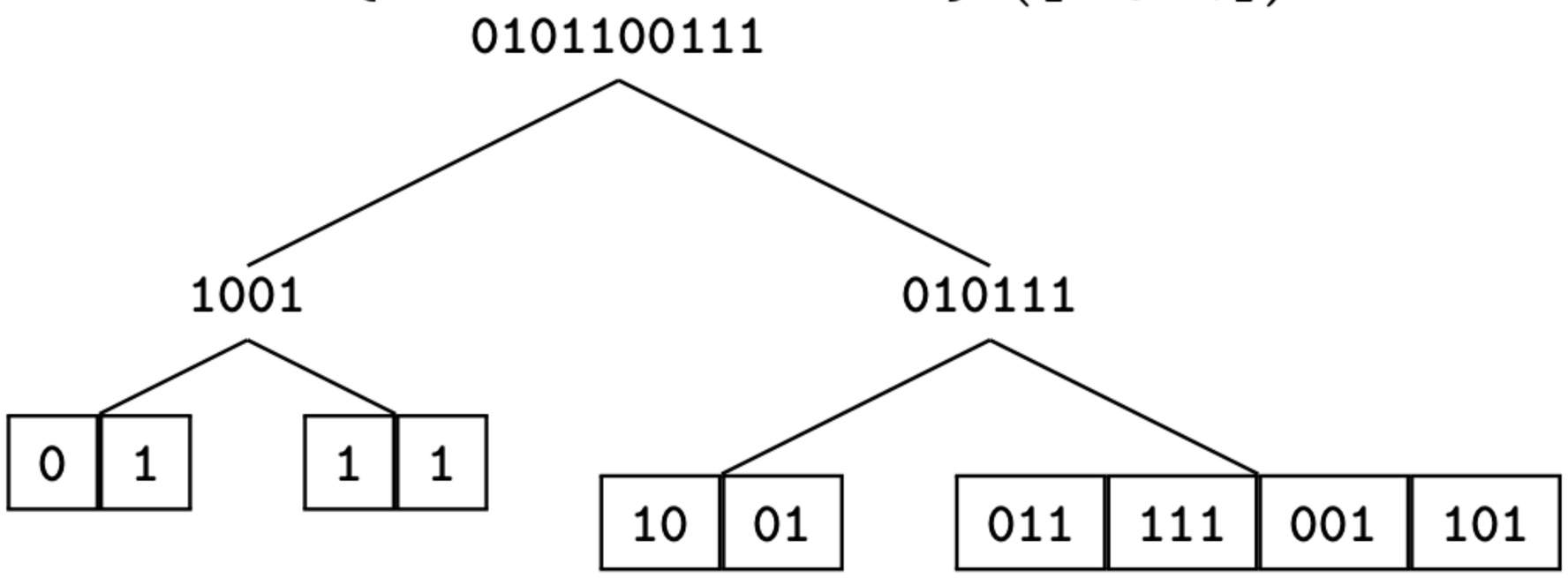
$2\lfloor \log x \rfloor + 1$ bits long

https://www.researchgate.net/profile/M-Kuelekci/publication/262938266_Enhanced_Variable-Length_Codes_Improved_Compression_with_Efficient_Random_Access/links/00b7d5396b8b1862c4000000/Enhanced-Variable-Length-Codes-Improved-Compression-with-Efficient-Random-Access.pdf

Compact Integer Codes with WT

$$X=\{3,6,0,11,5,1,3,15,9,13\}$$

$$\text{Labels\_EliasW}=\{1,2,0,3,2,0,1,3,3,3\}\ (\lfloor \log x_i \rfloor)$$



0	0
1	1
3	**1**1
5	**1**01
6	**1**10
9	**1**001
11	**1**011
13	**1**101
15	**1**111

https://www.researchgate.net/profile/M-Kuelekci/publication/262938266_Enhanced_Variable-Length_Codes_Improved_Compression_with_Efficient_Random_Access/links/00b7d5396b8b1862c4000000/Enhanced-Variable-Length-Codes-Improved-Compression-with-Efficient-Random-Access.pdf

Reading assignment

- Range quantile queries https://arxiv.org/pdf/0903.4726.pdf
- Integer representations with R/S and Wavelet Trees https://ieeexplore.ieee.org/abstract/document/6824444