CSL 356 Aug 8

<u>Radix sort</u>: We can sort integers in the range $[1, n^c]$ for any constant c in $O(c \cdot n)$ steps.

Recall that in phase we were using bucket sort/countsort that takes time $O(m + n)$

range of the values

elements

For the above result, we chose $m = n$ (by making the radix $= n$)

$$b_1\ b_2\ b_3 \ \ldots \ \overbrace{- \ - \ \ b_k}^{\log n}$$

range $= [1..n]$

range $(1..m_1)$ phase 1

" $(1..m_2)$ phase 2

$$\sum_i (m_i + n)$$

range of values in phase i

LSB $\to$ MSB

MSB $\to$ LSB

We have n strings $s_1, s_2, \ldots s_n$

$$\sum |s_i| = N$$

input size

$|s_i| = l_i$

$\max_i |s_i| = L$

at, ate, net, class

1 2 4 3

Σ : alph

Comparison : $O(n \log n \cdot L)$

$(1,a)^1 (2,t)^1$ a t 1

$(1,a)^2 (2,t)^2 (3,e)^2$ a t e 2 $\not{b}$

n e t 3

c l a s s 4

1	2	3	4	5
a	t	$\not{b}$	$\not{b}$	$\not{b}$
a	t	e	$\not{b}$	$\not{b}$
n	e	t	$\not{b}$	$\not{b}$
c	l	a	s	s

n

$\leftarrow L \rightarrow$

$O(L \cdot (n + |\Sigma|))$

$O(L \cdot n)$

$|\Sigma| < n$

How good is $O(L\,n)$ n $O((N-n)n)$

Input size : N

$$\sum l_i = N$$

how do make it large to get a sense of worst case bound?

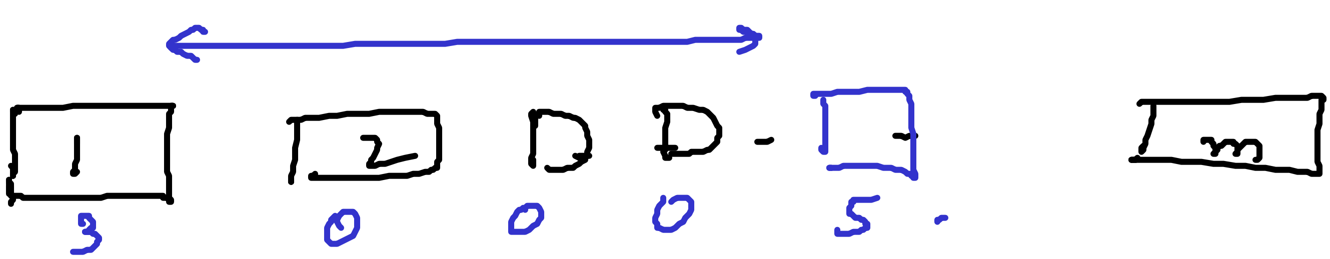
By choosing one long string and making other strings very short, say length 1

$$\Rightarrow \quad L + n - 1 = N$$

$$L = \underline{N - n}$$

at $n = \frac{N}{2}$ Running-time is $O(N^2)$

$$x_1 \quad x_2 \quad x_3 \quad \ldots \quad x_n$$



Objective : To collect all elements corresponding to a position together. And within each, we want them to be ordered according the non-empty buckets

Soln : Use radix sort on the N tuples.

Running time: $\underbrace{O(N+|\Sigma|)}_{1^{st}\text{ phase}} + \underbrace{O(N+L)}_{2^{nd}\text{ phase}}$

Overall $O(N+\cancel{L}+|\Sigma|)$

$L<N$

We have all the required information to run bucket sort for each column in time proportional to the number of non blank symbols

If the no. of non-blank symbols in column i is n_i, then it will take us $O(n_i)$ steps to run phase i of radix sort.

The time for phase i

$= O(m_i + n_i) = O(n_i)$

# of non-empty buckets

$m_i \leq n_i$

Total time $= O(\sum_i n_i) = O(N)$ for radix sort. ✓

Incl preprocessing $O(N + |\Sigma|)$

Issues to be resolved (on your own)

1. How Do we avoid the cost of moving strings (long ones) into the bucket during the bucket sort of each phase?

2. In round i, how do we introduce the new string of lengths $L - i$