מיונים
Sorts

1

מה נלמד?

- מיון מהיר Quicksort
- חסם תחתון על זמן ריצה של מיון מבוסס השוואות
- מיונים בזמן ליניארי
  - Counting sort
  - Radix Sort

2

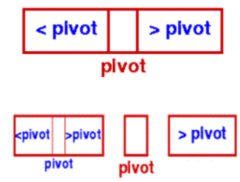
מיון מהיר
Quicksort

- פותח ב-1960 על ידי C.A.R.Hoare
- פעול לפי אסטרטגיית הפרד ומשול
- זמן הריצה במקרה הגרוע ביותר: $O(n^2)$
- זמן ריצה צפוי: $O(n \log n)$
- קבועים המסתתרים ב- $O(n \log n)$ קטנים
- ממיין במקום (in-place)

3

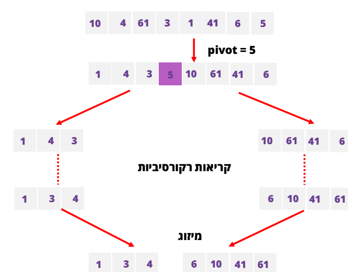
3

Quicksort
הרעיון



4

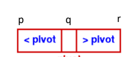
QuickSort



5

Quicksort

- To sort the sub-array A[p . . r]:
  - **Divide (חלק)**:
    - Partition A[p . . r] into A[p . . q − 1] and A[q + 1 . . r], such that
      - each element in A[p . . q − 1] is ≤ A[q] and
      - A[q] is ≤ each element in A[q + 1 . . r].
  - **Conquer (משול)**:
    - Sort the two sub-arrays by recursive calls to Quicksort.
  - **Combine (צרף)**:
    - No work is needed to combine the sub-arrays, because they are sorted in place.

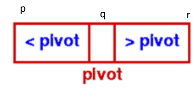


6

6

Quicksort

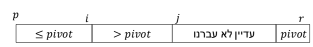
Quicksort (A, p, r)
  if $(p < r)$ **then**
    q = **partition** (A, p, r)
    Quicksort $(A, p, q - 1)$
    Quicksort $(A, q + 1, r)$



- **Initial call** is Quicksort $(A, 1, n)$.

7

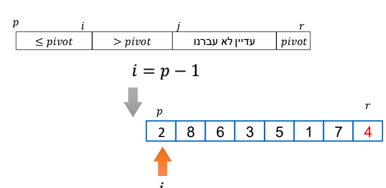
Partition



- $A[r]$ = pivot
- All entries in $A[p\,..\,i\,]$ are ≤ pivot
- All entries in $A[i + 1\,..\,j - 1]$ are > pivot
- All entries in $A[j\,..\,r - 1]$ are not yet examined.

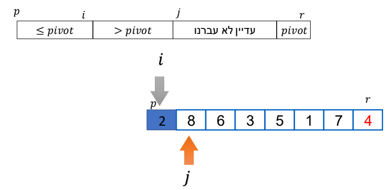
8

Partition



9

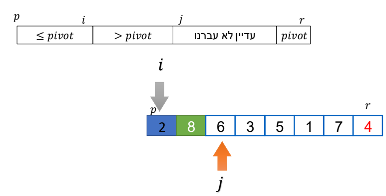
Partition



10

10

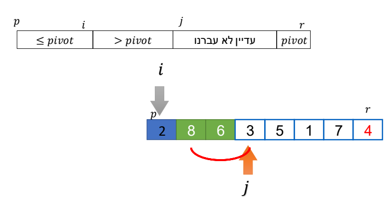
Partition



11

11

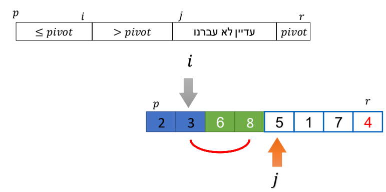
Partition



12

12

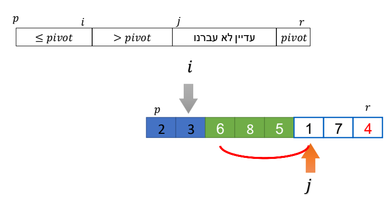
Partition



13

13

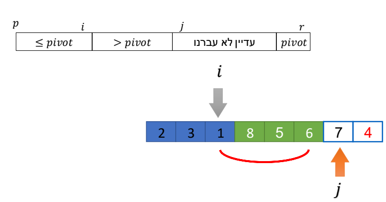
Partition



14

14

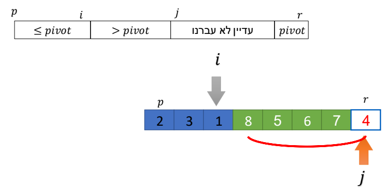
Partition



15

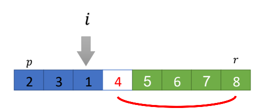
15

Partition



16

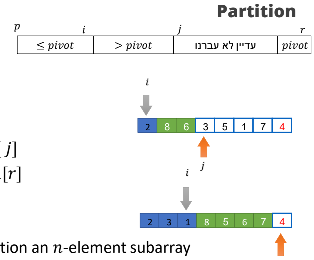
Partition



17

Partition

partition $(A, p, r\ )$
$\quad i\ =\ p\ -\ 1$
$\quad$ **for** $(j\ =\ p$ **to** $r\ -\ 1)$
$\quad\quad$ **if** $(A[j]\ \leq\ A[r])$
$\quad\quad\quad i\ =\ i\ +\ 1$
$\quad\quad\quad$ **exchange** $A[i]$ **with** $A[\ j]$
$\quad$ **exchange** $A[i\ +\ 1]$ **with** $A[r]$
$\quad$ **return** $i\ +\ 1$



- **Complexity**: $\Theta(n)$ to partition an n-element subarray

18

זמן ריצה של מיון מהיר
Performance

- The running time of Quicksort depends on the partitioning of the sub-arrays:
  - If the sub-arrays are balanced, then Quicksort can run as fast as Mergesort.
  - If they are unbalanced, then Quicksort can run as slowly as Insertion sort.

19

19

מקרה הגרוע ביותר
Worst Case

- Occurs when the sub-arrays are **completely unbalanced** every time.
  - When Quicksort takes a sorted array as input.
- Have 0 elements in one sub-array and $n-1$ elements in the other sub-array.
- Get the recurrence

$$\begin{aligned} T(n) &= T(n-1) + T(0) + \Theta(n) \\ &= T(n-1) + \Theta(n) \\ &= \Theta(n^2) . \end{aligned}$$

- Same running time as insertion sort.

20

20

מקרה הטוב ביותר
Best Case

- Occurs when the sub-arrays are **completely balanced** every time.
- Each sub-array has $\approx n/2$ elements.
- Get the recurrence

$$\begin{aligned} T(n) &= 2T(n/2) + \Theta(n) \\ &= \Theta(n \lg n) . \end{aligned}$$

21

21

זמן ריצה צפוי
expected running time

- If the pivot is the k'th element, the runtime is $T(n)=T(k)+T(n-1-k)+cn$
- The probability that the pivot is the k'th element is $1/n$
- Therefore, the expected runtime can be expressed as:

$$T(n)=cn+\frac{1}{n}\sum_{k=0}^{n-1}\big(T(k)+T(n-1-k)\big)$$

22

22

זמן ריצה צפוי
expected running time

$$T(n)=cn+\frac{1}{n}\sum_{k=0}^{n-1}\big(T(k)+T(n-1-k)\big)$$

$$nT(n)=cn^2+2\sum_{k=0}^{n-1}T(k)(*)$$

$$(n-1)T(n-1)=c(n-1)^2+2\sum_{k=0}^{n-2}T(k)(**)$$

$$nT(n)-(n-1)T(n-1)=c(2n-1)+2T(n-1)$$

$$nT(n)=c(2n-1)+(n+1)T(n-1)$$

$$\frac{T(n)}{n+1}=\frac{c(2n-1)}{n(n+1)}+\frac{T(n-1)}{n}\le\frac{2nc}{n(n+1)}+\frac{T(n-1)}{n}=\frac{2c}{n+1}+\frac{T(n-1)}{n}$$

$$\frac{T(n)}{n+1}\le\frac{2c}{n+1}+\frac{T(n-1)}{n}\le\frac{2c}{n+1}+\frac{2c}{n}+\frac{T(n-2)}{n-1}\le\ldots\le 2c\sum_{i=3}^{n+1}\frac{1}{i}+\frac{T(1)}{2}$$

$$\boxed{T(n)=O(n\log n)}$$

23

23

זמן ריצה של מיון מהיר
סיכום

- זמן הריצה במקרה הגרוע ביותר: $O(n^2)$
- זמן ריצה צפוי: $O(n\log n)$
- קבועים המסתתרים ב- $O(n\log n)$ קטנים
- ממיין במקום (in-place)

24

MergeSort	$O(n\log n)$
HeapSort	$O(n\log n)$
Insertion Sort	$O(n^2)$
QuickSort	$O(n\log n)$ **בהסתברות גבוהה**

?

האם קיים אלגוריתם מיון מבוסס השוואות שזמן ריצתו במקרה הגרוע קטן מ- $O(n\log n)$?

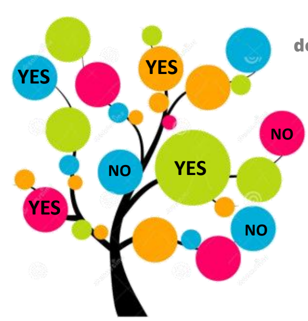
25

מיונים מבוססים השוואה
Comparisons based sorts

- מיון מבוסס השוואות: הוא מיון שבו מידע על סדר האיברים מתקבל אך ורק ע"י השוואת שני איברים.
- כל המיונים שראינו עד עכשיו הם מבוססי השוואות.
- סיבוכיות הזמן של כל מיון מבוסס השוואות היא $\Omega(n\log n)$.

26

עץ החלטה – decision tree



- ניצג אלגוריתם מבוסס השוואות בעזרת עץ החלטה - **decision tree**

$$>, <, \geq, \leq$$

- בלי הגבלת הכלליות, נניח שכל המספרים שונים זה מזה, ונגביל את עצמינו להשוואה "$\leq$"

27

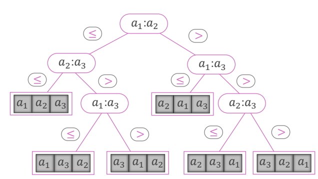
עץ החלטה – decision tree
מיון הכנסה של מערך בגודל n=3

a_1	a_2	a_3

28

עץ החלטה – decision tree
מיון הכנסה של מערך בגודל n=3

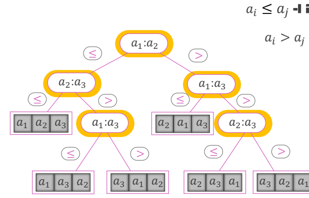
a_1	a_2	a_3



29

עץ החלטה – decision tree

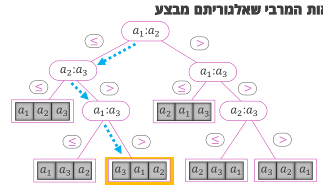
- כל צומת פנימית $a_i : a_j$ מסמנת השוואה בין a_i לבין a_j
- תת-העץ השמאלי מתאר החלטות במידה ו- $a_i \le a_j$
- תת-העץ הימני מתאר החלטות במידה ו- $a_i > a_j$
- כל עלה הוא תמורה של איברי המערך



30

עץ החלטה – decision tree

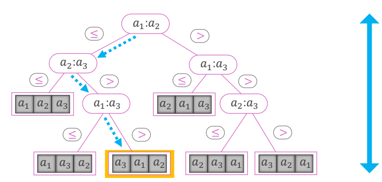
- כל מסלול משורש לעלה מתאר סידרת השוואות אפשרית שמבצע אלגוריתם על מערך הקלט בגודל n
- המסלול הכי ארוך מתאר מספר השוואות המרבי שאלגוריתם מבצע



31

עץ החלטה – decision tree

- המסלול הכי ארוך מתאר מספר השוואות המרבי שאלגוריתם מבצע
- גובה של עץ ההחלטה קובע חסם תחתון על זמן ריצה במקרה הגרוע של אלגוריתם מיון



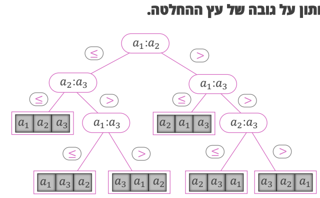
32

משפט: החסם התחתון של מיון מבוסס השוואות

כל מיון מבוסס השוואות על מערך בגודל n דורש $\Omega(n \log n)$ השוואות במקרה הגרוע.

הוכחה

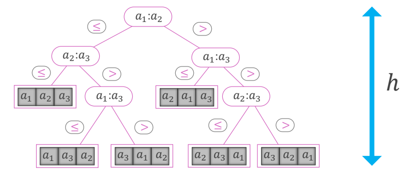
מהדיון הקודם יש למצוא חסם תחתון על גובה של עץ ההחלטה.



33

? מהו חסם תחתון על גובה של עץ ההחלטה

1. $\Omega(n \log n)$
2. $\Omega(n^2)$
3. $\Omega(n \log^2 n)$
4. $\Omega(n^{1.5} \log n)$



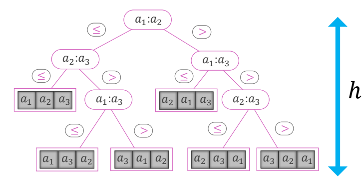
34

משפט החסם התחתון של מיון מבוסס השוואות – הוכחה

יהי T עץ החלטה בגובה h המתאר מיון מבוסס השוואות.

כל אחת מ $n!$ תמורות של איברי המערך חייבת להופיע באחד העלים.

מצד שני, מספר המירבי של עלים שיכול להיות בעץ בינארי בגובה h הוא 2^h.



מכאן, $n! \leq 2^h$.

$h \geq \log n!$

מכאן, $h = \Omega(n \log n)$.

מ.ש.ל.

36

<u>מסקנה</u>: זמן ריצה במקרה הגרוע של כל אלגוריתם מבוסס השוואות על קלט בגודל n חסום מלמטה ע"י $\Omega(n \log n)$.

? האם קיים איזשהו אלגוריתם למיון שזמן ריצתו במקרה הגרוע קטן יותר מ $O(n\log n)$

! לא – אם האלגוריתם משתמש בהשוואות בלבד
כן – אם אלגוריתם משתמש במידע נוסף על מערך הקלט

37

מיונים בזמן ליניארי

- ישנם מיונים לא מבוססי השוואות שסיבוכיות הזמן שלהם היא ליניארית כלומר $O(n)$.
- מיונים אלה מניחים הנחות מסוימות על הקלט.
  - מיון מנייה counting sort
  - מיון בסיס Radix Sort

38

מיון מניה
Counting Sort

- הנחה: איברי הקלט הם מספרים שלמים בתחום מ- 0 עד k, עבור k שלם חיובי
- <u>רעיון</u>: לכל איבר x בקלט סופרים את כמות האיברים שקטנים או שווים לו (כולל x עצמו).

A	2	5	3	0	2	3	0	3

A לאחר מיון

39

מיון מנייה
שלבי האלגוריתם

1. שלב המנייה
2. שלב הצבירה
3. בניית מערך הפלט

40


מיון מניה דוגמת הרצה
A
2 5 3 0 2 3 0 3
k = 5
C
0 0 0 0 0 0
שלב I - שלב מנייה
עבור כל איבר i, נספור כמה איברים שווים ל i יש ב A
C[i] – כמות האיברים ששווים ל i

41

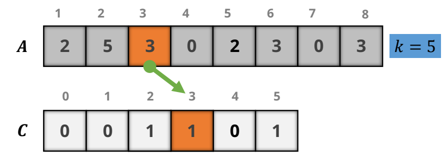

מיון מניה דוגמת הרצה
A
2 5 3 0 2 3 0 3
k = 5
C
0 0 1 0 0 0
שלב I - שלב מנייה
עבור כל איבר i, נספור כמה איברים שווים ל i יש ב A
C[i] – כמות האיברים ששווים ל i

42


מיון מניה דוגמת הרצה
A
2 5 3 0 2 3 0 3
k = 5
C
0 0 1 0 0 1
שלב I - שלב מנייה
עבור כל איבר i, נספור כמה איברים שווים ל i יש ב A
C[i] – כמות האיברים ששווים ל i

43

מיון מניה דוגמת הרצה

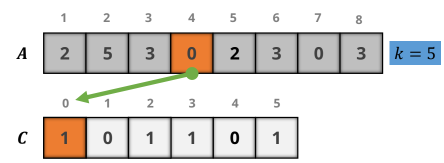


שלב I - שלב מנייה

- עבור כל איבר i, נספור כמה איברים שווים ל i יש ב A
- $C[i]$ – כמות האיברים ששווים ל i

44

מיון מניה דוגמת הרצה

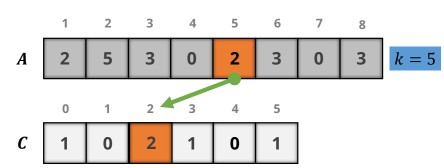


שלב I - שלב מנייה

- עבור כל איבר i, נספור כמה איברים שווים ל i יש ב A
- $C[i]$ – כמות האיברים ששווים ל i

45

מיון מניה דוגמת הרצה



שלב I - שלב מנייה

- עבור כל איבר i, נספור כמה איברים שווים ל i יש ב A
- $C[i]$ – כמות האיברים ששווים ל i

46

מיון מניה דוגמת הרצה

	1	2	3	4	5	6	7	8	
A	2	5	3	0	2	3	0	3	$k=5$

	0	1	2	3	4	5
C	2	0	2	3	0	1

שלב I - שלב מנייה

- עבור כל איבר i, נספור כמה איברים שווים ל i יש ב A
- $C[i]$ – כמות האיברים ששווים ל i

47

מיון מניה דוגמת הרצה

	1	2	3	4	5	6	7	0	
A	2	5	3	0	2	3	0	3	$k=5$

	0	1	2	3	4	5
C	2	0	2	3	0	1

≤ 0

שלב II - שלב צבירה

- נעבור על C משמאל לימין, ונסכום $C[i] \leftarrow C[i] + C[i-1]$
- $C[i]$ – כמות האיברים שקטנים או שווים ל i

48

מיון מניה דוגמת הרצה

	1	2	3	4	5	6	7	8	
A	2	5	3	0	2	3	0	3	$k=5$

	0	1	2	3	4	5
C	2	2	2	3	0	1

≤ 1

שלב II - שלב צבירה

- נעבור על C משמאל לימין, ונסכום $C[i] \leftarrow C[i] + C[i-1]$
- $C[i]$ – כמות האיברים שקטנים או שווים ל i

49

מיון מניה דוגמת הרצה

	1	2	3	4	5	6	7	8	
A	2	5	3	0	2	3	0	3	$k = 5$

	0	1	2	3	4	5
C	2	2	4	3	0	1

≤ 2

שלב II - שלב צבירה

- נעבור על C משמאל לימין, ונסכום $C[i] \leftarrow C[i] + C[i-1]$
- $C[i]$ – כמות האיברים שקטנים או שווים ל i

50

מיון מניה דוגמת הרצה

	1	2	3	4	5	6	7	8	
A	2	5	3	0	2	3	0	3	$k = 5$

	0	1	2	3	4	5
C	2	2	4	7	0	1

≤ 3

שלב II - שלב צבירה

- נעבור על C משמאל לימין, ונסכום $C[i] \leftarrow C[i] + C[i-1]$
- $C[i]$ – כמות האיברים שקטנים או שווים ל i

51

מיון מניה דוגמת הרצה

	1	2	3	4	5	6	7	8	
A	2	5	3	0	2	3	0	3	$k = 5$

	0	1	2	3	4	5
C	2	2	4	7	7	1

≤ 4

שלב II - שלב צבירה

- נעבור על C משמאל לימין, ונסכום $C[i] \leftarrow C[i] + C[i-1]$
- $C[i]$ – כמות האיברים שקטנים או שווים ל i

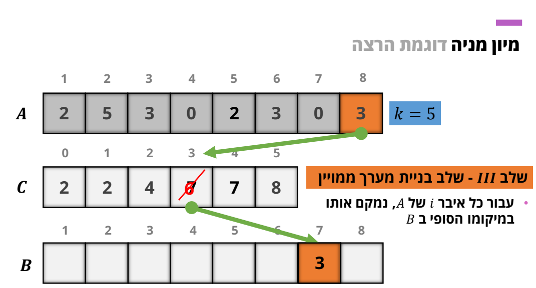
52


מיון מניה דוגמת הרצה
1 2 3 4 5 6 7 8
A
2 5 3 0 2 3 0 3
k = 5
0 1 2 3 4 5
C
2 2 4 7 7 8
שלב II - שלב צבירה
נעבור על C משמאל לימין, ונסכום C[i] ← C[i] + C[i − 1]
C[i] – כמות האיברים שקטנים או שווים ל i

53


מיון מניה דוגמת הרצה
1 2 3 4 5 6 7 8
A
2 5 3 0 2 3 0 3
k = 5
0 1 2 3 4 5
C
2 2 4 7 7 8
שלב III - שלב בניית מערך ממויין
עבור כל איבר i של A, נמקם אותו במיקומו הסופי ב B
1 2 3 4 5 6 7 8
B

54


מיון מניה דוגמת הרצה
1 2 3 4 5 6 7 8
A
2 5 3 0 2 3 0 3
k = 5
0 1 2 3 4 5
C
2 2 4 6 7 8
שלב III - שלב בניית מערך ממויין
עבור כל איבר i של A, נמקם אותו במיקומו הסופי ב B
1 2 3 4 5 6 7 8
B
3

55


מיון מניה דוגמת הרצה
A
2 5 3 0 2 3 0 3
k = 5
C
2 2 4 6 7 8
B
0 3
שלב III - שלב בניית מערך ממויין
עבור כל איבר i של A, נמקם אותו
במיקומו הסופי ב B

56


מיון מניה דוגמת הרצה
A
2 5 3 0 2 3 0 3
k = 5
C
1 2 4 6 7 8
B
0 3 3
שלב III - שלב בניית מערך ממויין
עבור כל איבר i של A, נמקם אותו
במיקומו הסופי ב B

57


מיון מניה דוגמת הרצה
A
2 5 3 0 2 3 0 3
k = 5
C
1 2 4 5 7 8
B
0 2 3 3
שלב III - שלב בניית מערך ממויין
עבור כל איבר i של A, נמקם אותו
במיקומו הסופי ב B

58


מיון מניה דוגמת הרצה
$k = 5$
שלב III - שלב בניית מערך ממוין
עבור כל איבר i של A, נמקם אותו במיקומו הסופי ב B

59


מיון מניה דוגמת הרצה
$k = 5$
שלב III - שלב בניית מערך ממוין
עבור כל איבר i של A, נמקם אותו במיקומו הסופי ב B

60


מיון מניה דוגמת הרצה
$k = 5$
שלב III - שלב בניית מערך ממוין
עבור כל איבר i של A, נמקם אותו במיקומו הסופי ב B

61

מיון מניה דוגמת הרצה

	1	2	3	4	5	6	7	8
A	2	5	3	0	2	3	0	3

$k = 5$

	0	1	2	3	4	5
C	0	2	$\cancel{3}$	4	7	7

	1	2	3	4	5	6	7	8
B	0	0	2	2	3	3	3	5

שלב III - שלב בניית מערך ממויין

- עבור כל איבר i של A, נמקם אותו במיקומו הסופי ב B

62

מיון מניה דוגמת הרצה

	1	2	3	4	5	6	7	8
A	2	5	3	0	2	3	0	3

$k = 5$

	0	1	2	3	4	5
C	0	2	2	4	7	7

	1	2	3	4	5	6	7	8
B	0	0	2	2	3	3	3	5

שלב III - שלב בניית מערך ממויין

- עבור כל איבר i של A, נמקם אותו במיקומו הסופי ב B

63

מיון מניה
זמן ריצה

	1	2	3	4	5	6	7	8
A	2	5	3	0	2	3	0	3

	0	1	2	3	4	5
C	0	0	0	0	0	0

Counting Sort(A, B, k)

1. **for** $(i \leftarrow 0$ **to** $k)$
2. $\quad C[i] \leftarrow 0$

64

מיון מניה
זמן ריצה

$Counting\ Sort(A, B, k)$

```
for (i ← 0 to k)
    C[i] ← 0
for (i ← 1 to A.length)
    C[A[i]] ← C[A[i]] + 1
```

	1	2	3	4	5	6	7	8
A	2	5	3	0	2	3	0	3

	0	1	2	3	4	5
C	2	0	2	3	0	1

65

מיון מניה
זמן ריצה

$Counting\ Sort(A, B, k)$

```
for (i ← 0 to k)
    C[i] ← 0
for (i ← 1 to A.length)
    C[A[i]] ← C[A[i]] + 1
for (i ← 1 to k)
    C[i] ← C[i] + C[i − 1]
```

	1	2	3	4	5	6	7	8
A	2	5	3	0	2	3	0	3

	0	1	2	3	4	5
C	2	2	4	7	7	8

66

מיון מניה
זמן ריצה

$Counting\ Sort(A, B, k)$

```
for (i ← 0 to k)
    C[i] ← 0
for (i ← 1 to A.length)
    C[A[i]] ← C[A[i]] + 1
for (i ← 1 to k)
    C[i] ← C[i] + C[i − 1]
for (j ← A.length downto 1)
    B[C[A[j]]] ← A[j]
    C[A[j]] − −
```

	1	2	3	4	5	6	7	8
A	2	5	3	0	2	3	0	3

	0	1	2	3	4	5
C	2	2	4	7	7	8

	1	2	3	4	5	6	7	8
B	0	0	2	2	3	3	3	5

67

מיון מניה
זמן ריצה

1	2	3	4	5	6	7	8
0	0	2	2	3	3	3	5

```
Counting Sort(A, B, k)
for (i ← 0 to k)
    C[i] ← 0
for (i ← 1 to A.length)
    C[A[i]] ← C[A[i]] + 1
for (i ← 1 to k)
    C[i] ← C[i] + C[i − 1]
for (j ← A.length downto 1)
    B[C[A[j]]] ← A[j]
    C[A[j]] − −
return B
```

68

מיון מניה
זמן ריצה

```
Counting Sort(A, B, k)
for (i ← 0 to k)                      O(k)
    C[i] ← 0
for (i ← 1 to A.length)               O(n)
    C[A[i]] ← C[A[i]] + 1
for (i ← 1 to k)                      O(k)
    C[i] ← C[i] + C[i − 1]
for (j ← A.length downto 1)
    B[C[A[j]]] ← A[j]                 O(n)
    C[A[j]] − −
return B
```

זמן ריצה כולל הינו $O(n+k)$

69

? אם בבניית מערך הפלט היינו עוברים על A מהתחלה לסוף (ולא מסוף להתחלה), האם המיון היה נכון ?

	1	2	3	4	5	6	7	8
A	2	5	3	0	2	3	0	3

	0	1	2	3	4	5
C	2	2	4	7	7	8

	1	2	3	4	5	6	7	8
B	0	0	2	2	3	3	3	5

1. כן, אין חשיבות לכיוון המעבר
2. לא, רק מעבר מסוף להתחלה היה נותן מיון תקין
3. אפשר לעבור מהתחלה לסוף, אך היינו מאבדים תכונת מיון חשובה

70

אם בבניית מערך הפלט היינו עוברים על A מהתחלה לסוף (ולא מסוף להתחלה), האם המיון היה נכון ?

	1	2	3	4	5	6	7	8
A	2_1	5	3_1	0_1	2_2	3_2	0_2	3_3

	1	2	3	4	5	6	7	8
B	0_2	0_1	2_2	2_1	3_3	3_2	3_1	5

	1	2	3	4	5	6	7	8
B	0_1	0_2	2_1	2_2	3_1	3_2	3_3	5

71

מיון יציב הגדרה

שמירה של הסדר המקורי בין ערכים זהים נקראת **תכונת היציבות** – stability, **ומיון שיש לו את התכונה הזו נקרא מיון יציב** – stable sort.

	1	2	3	4	5	6	7	8
A	2_1	5	3_1	0_1	2_2	3_2	0_2	3_3

	1	2	3	4	5	6	7	8
B	0_1	0_2	2_1	2_2	3_1	3_2	3_3	5

72

מיון בסיס
Radix Sort

- הנחה: מספר הקלט הם שלמים בני d ספרות
- דוגמה:

73

למה חשוב להשתמש במיון יציב למיון ספרות?
מה עלול לקרות אם היינו משתמשים במיון לא יציב?

.1	.2	.3	.4
אפשר גם מיון לא יציב, זה לא משנה	מיון לא יציב היה עלול להחזיר תוצאת מיון שגויה	מיון לא יציב תמיד היה מחזיר תוצאת מיון שגויה	מיון יציב חשוב לנו רק כדי לקבל זמן ריצה טוב

74

מיון בסיס
Radix Sort

RadixSort (A, d)

for $(i = 1 \text{ to } d)$ **do**

use a **stable sort** to sort array A on digit i

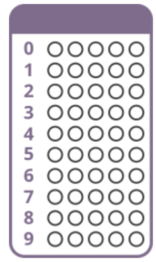
75

Radix Sort
זמן ריצה

RadixSort (A, d)
for $(i = 1 \text{ to } d)$ do
use a stable sort to sort array A on digit i

- Assume that we use counting sort as the intermediate sort.
- $\Theta(n + k)$ per pass (digits in range $0, \ldots, k$).
- d passes.
- $\Theta(d(n + k))$ total.
- If $k = O(n)$ and $d = O(1)$, then $T(n) = \Theta(n)$.

76



77

סיכום

78