

Metode de sortare

- pregătire admitere -

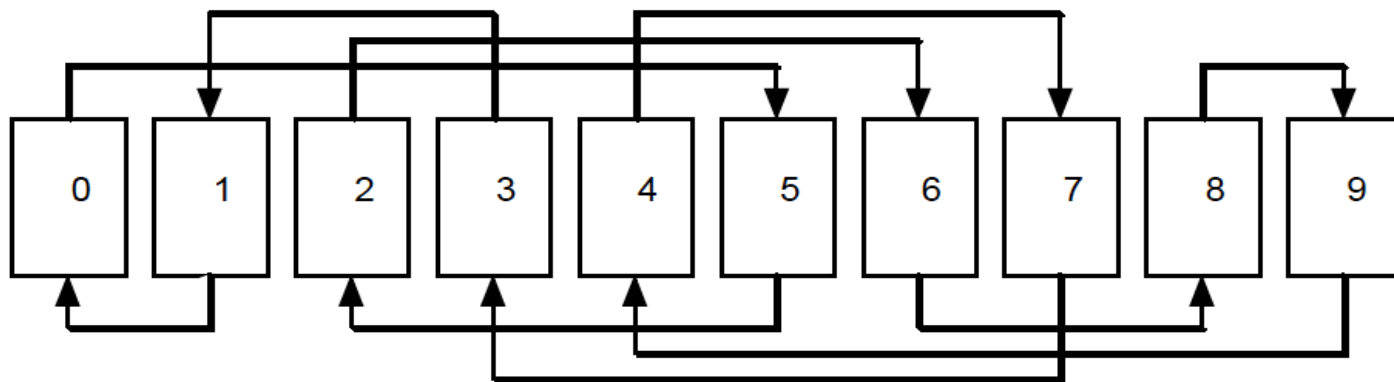
Universitatea din București

Cuprins

- **Problema sortării**
- Algoritmul de sortare prin interschimbare (Bubble sort)
- Algoritmul de sortare prin inserție (Insertion sort)
- Algoritmul de sortare prin selecție (Selection sort)
- Algoritmul de sortare prin interclasare (Mergesort)
- Algoritmul de sortare rapidă (Quicksort)

Problema sortării

- Fie un șir de “entități”: e.g. numere, caractere, înregistrări ale unei baze de date, etc. Fiecare entitate deține o caracteristică numită **cheie de sortare**.



http://www.whymath.org/Reading_Room_Material/ian_stewart/shuffle/shuffle.html

- Există o **relație de ordine** ce se definește peste mulțimea valorilor posibile ale cheii de sortare.
- Sortarea șirului = dispunerea elementelor șirului astfel încât valorile cheilor de sortare să se afle în ordine crescătoare (sau descrescătoare)

Problema sortării - Exemple

- Elemente numerice: [1, 5, 2, 3, 6, 10, 22, 4, 44]. În acest caz cheia de sortare este reprezentată de însuși elementul șirului.
- Elemente definite de înregistrări cu mai multe câmpuri (nume și vârstă):
(Andrei, 29), (Paul, 29), (Tiberiu, 10), (Corina, 25), (Claudiu, 19).

Două posibile chei de sortare pentru fiecare element:

- Nume
- Vârstă

Sortare după nume:

(Andrei, 29), (Claudiu, 19), (Corina, 25), (Paul, 29), (Tiberiu, 10).

Sortare după vârstă:

(Tiberiu, 10), (Claudiu, 19), (Corina, 25), (Andrei, 29), (Paul, 29).

Problema sortării

- Fie șirul:

$$x[1], \quad x[2], \quad \dots x[n].$$

Dacă notăm cheia elementului x_i cu K_i , atunci sortarea șirului presupune determinarea unei permutări $p(1), p(2), \dots, p(n)$ astfel încât:

$$K_{p(1)} \leq K_{p(2)} \leq \dots \leq K_{p(n)}.$$

- Pentru simplitate, considerăm că cheia de sortare K_i este reprezentată de întregul element x_i .

Proprietăți ale metodelor de sortare

1. **Eficiența.** Orice metodă de sortare este caracterizată de volumul de resurse (timp de execuție) necesitat pentru rezolvarea problemei.
 - Număr de comparații (de chei)
 - Număr de deplasări (de date)
2. **Stabilitate.** O metodă de sortare este stabilă dacă ordinea relativă a elementelor cu aceeași cheie de sortare rămâne neschimbată.

Aranjament inițial:

(Andrei, 29), (Paul, 29), (Tiberiu, 10), (Corina, 25), (Claudiu, 19).

Sortare **stabilă**:

(Tiberiu, 10), (Claudiu, 19), (Corina, 25), (Andrei, 29), (Paul, 29).

Sortare **instabilă**:

(Tiberiu, 10), (Claudiu, 19), (Corina, 25), (Paul, 29), (Andrei, 29).

3. **Simplitate.** O metodă de sortare este caracterizată de nivelul de complexitate al implementării sale în practică.

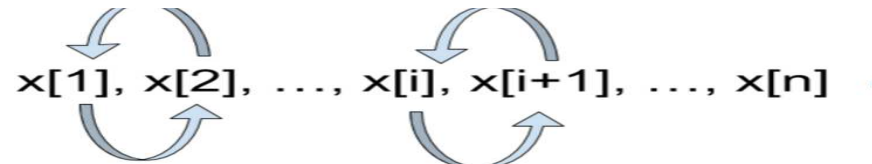
Cuprins

- Problema sortării
- **Algoritmul de sortare prin interschimbare (Bubble sort)**
- Algoritmul de sortare prin inserție (Insertion sort)
- Algoritmul de sortare prin selecție (Selection sort)
- Algoritmul de sortare prin interclasare (Mergesort)
- Algoritmul de sortare rapidă (Quicksort)

Sortare prin interschimbare (Bubble sort)

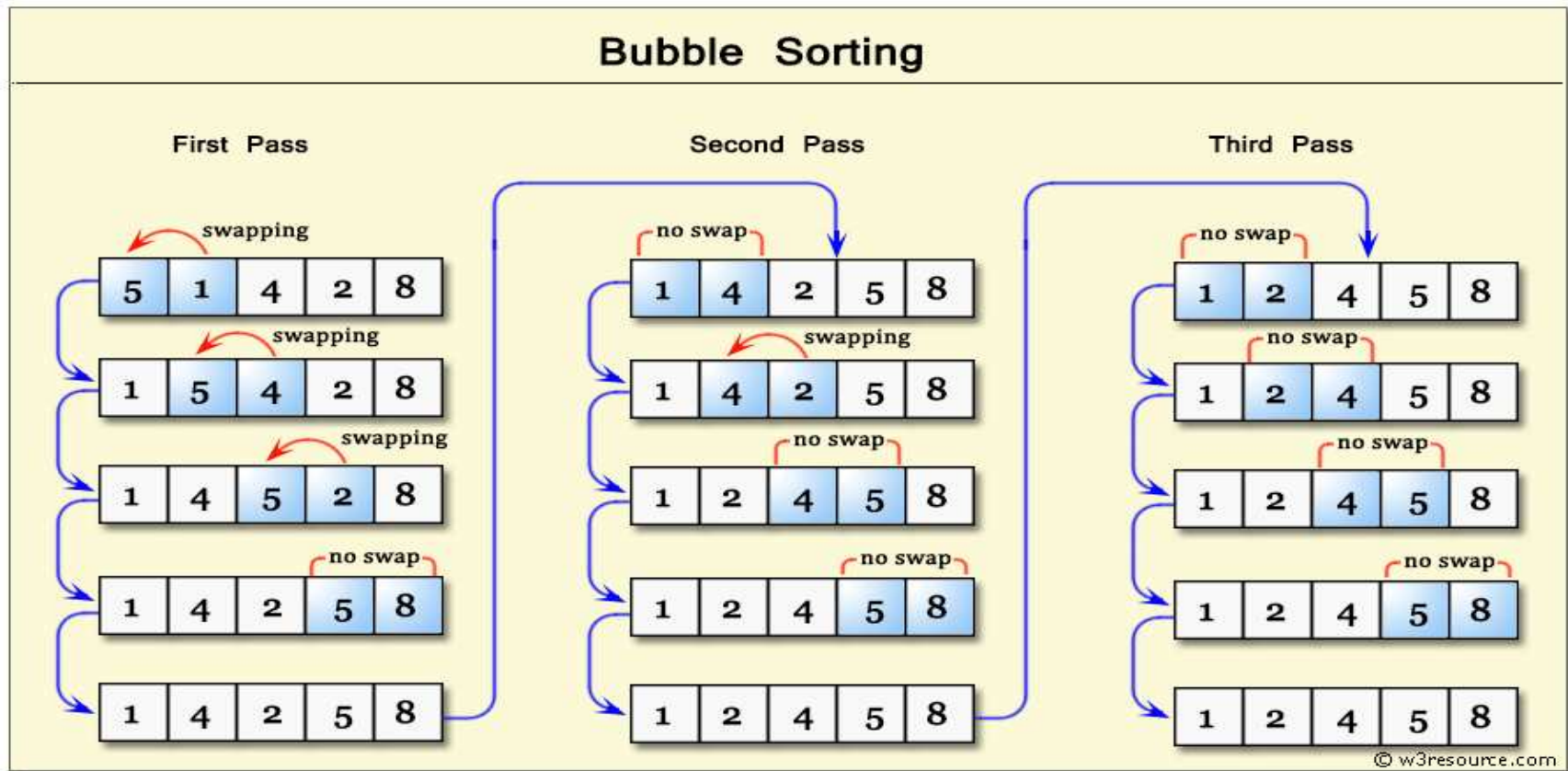
Idei generale:

- Considerăm șirul: $x[1], x[2], \dots, x[n]$.
- Se parcurge secvențial șirul (element cu element). Se compară elementul curent cu următorul și, în cazul în care se află în ordinea nepotrivită, se interschimbă.



- Se efectuează parcurgeri până șirul nu mai necesită interschimbări

Sortare prin interschimbare - Exemplu



Sortare prin interschimbare

Algoritm **Sort-Interschimbare**(array x):

```
1. repeat
    1. for  $i = 2 \dots n$ 
        1. if ( $x[i] < x[i - 1]$ )      // dacă  $x[i]$  și  $x[i - 1]$  nu sunt ordonate
            1.  $x[i - 1] \leftrightarrow x[i]$     // se interschimbă  $x[i]$  cu  $x[i - 1]$ 
            2.  $swap = \text{true}$ 
        endif
    endfor
until not swap    // repetă până nu mai sunt necesare interschimbări
1. return  $x$ 
```

Sortare prin interschimbare - Eficiență

Algoritm **Sort-Interschimbare**(array x):

```
1. repeat
    1. for  $i = 2 \dots n$ 
        1. if  $(x[i] < x[i - 1])$ 
            1.  $x[i - 1] \leftrightarrow x[i]$ 
            2.  $swap = \text{true}$ 
        endif
    endfor
until not swap
2. return  $x$ 
```

Operații:

- **Comparații** $T_C(n)$
- **Deplasări elemente** $T_D(n)$

Pentru fiecare i : 1 comparații

Pentru toate valorile i :

$$\sum_{i=2}^n 1 = n - 1$$

Pentru toate parcurgerile:

$$n - 1 \leq T_C(n) \leq \frac{n(n - 1)}{2}$$

Sortare prin interschimbare - Eficiență

Algoritm **Sort-Interschimbare**(array x):

```
1. repeat
    1. for  $i = 2 \dots n$ 
        1. if  $(x[i] < x[i - 1])$ 
            1.  $x[i - 1] \leftrightarrow x[i]$ 
            2.  $swap = \text{true}$ 
        endif
    endfor
until not swap
2. return  $x$ 
```

Operații:

- **Comparații** $T_C(n)$
- **Deplasări elemente** $T_D(n)$

Pentru fiecare i : **Maxim 2**

Pentru toate valorile i :

$$0 \leq T_i \leq 2(n - i)$$

Pentru toate parcurgerile:

$$0 \leq T_D(n) \leq n(n - 1)$$

Sortare prin interschimbare - Eficiență

- Comparații: $n - 1 \leq T_C(n) \leq \frac{n(n+1)}{2}$
- Deplasări elemente: $0 \leq T_D(n) \leq n(n - 1)$

- Complexitate totală:

$$n - 1 \leq T(n) = T_C(n) + T_D(n) \leq \frac{n(n+1)}{2} + n(n-1) \approx \mathcal{O}(n^2)$$

- Cel mai favorabil caz: $1, 2, 3, \dots, n$; Cel mai nefavorabil: $n, n - 1, \dots, 1$
- Algoritmul de sortare prin interschimbare este **stabil**!
- Simplu de implementat, dar ineficient pentru șiruri de dimensiuni mari!

Cuprins

- Problema sortării
- Algoritmul de sortare prin interschimbare (Bubble sort)
- **Algoritmul de sortare prin inserție (Insertion sort)**
- Algoritmul de sortare prin selecție (Selection sort)
- Algoritmul de sortare prin interclasare (Mergesort)
- Algoritmul de sortare rapidă (Quicksort)

Sortare prin inserție (Insertion sort)

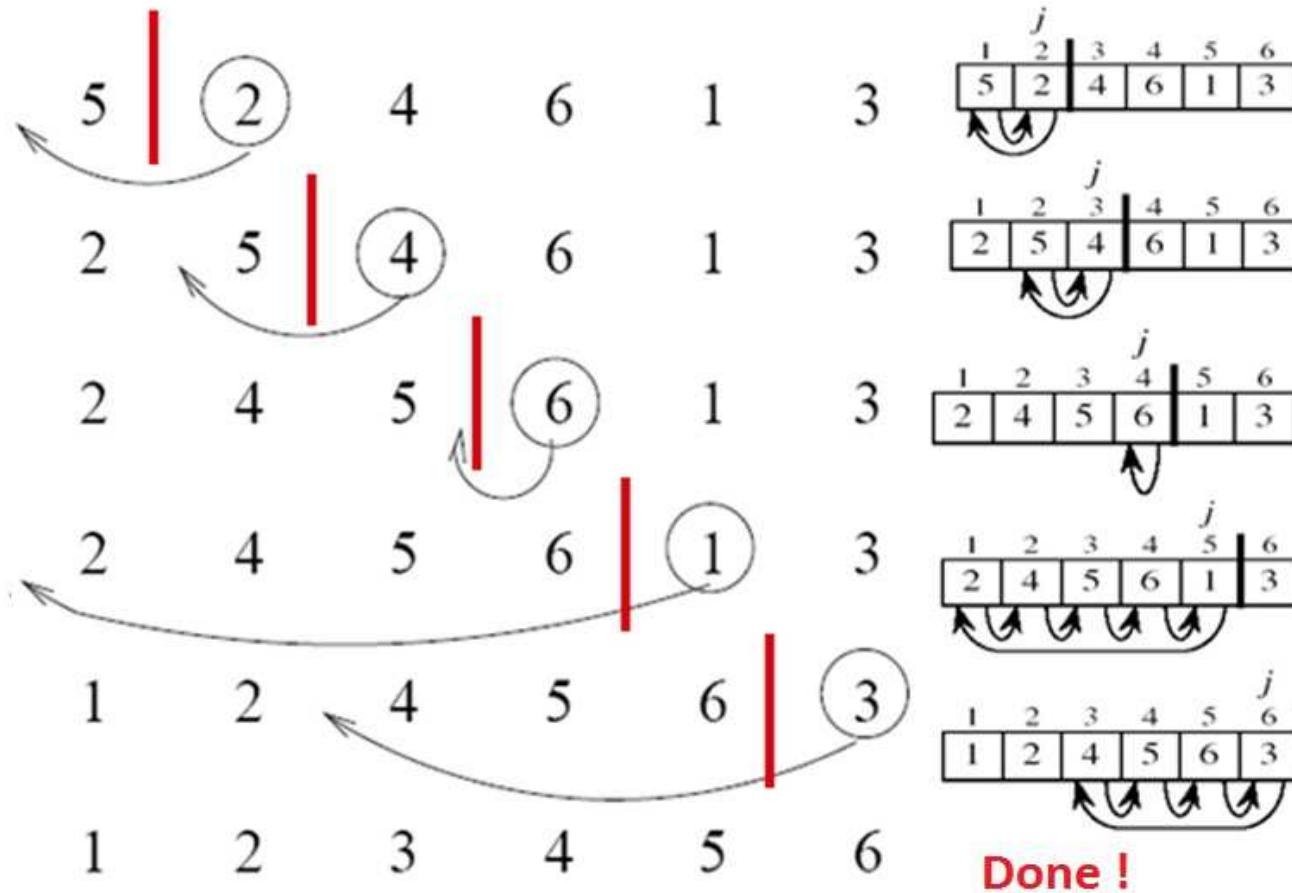
Idei generale:

- Considerăm șirul: $x[1], x[2], \dots, x[n]$.
- Se parcurge secvențial șirul (element cu element). Se inserează elementul curent ($x[i]$) în subșirul care îl precede ($x[1], x[2], \dots, x[i-1]$) astfel încât acesta să rămână ordonat:

$$x[1], x[2], \dots, x[i-1], x[i], x[i+1], \dots, x[n]$$

- Subșirul ce conține elementele deja sortate crește la fiecare iterație, astfel încât, după ce parcurgem toate elementele, secvența este sortată în întregime

Sortare prin inserție - Exemplu



<http://freefeast.info/general-it-articles/insertion-sort-pseudo-code-of-insertion-sort-insertion-sort-in-data-structure/>

Sortare prin inserție

Algoritm **Sort-Inserție**(array x):

1. **for** $i = 2 \cdots n$
 1. $\text{aux} \leftarrow x[i]$ // fixarea elementului i
 2. $j \leftarrow i - 1$
 3. **while** $(j \geq 1)$ and $(\text{aux} < x[j])$ // căutarea poziției
 1. $x[j + 1] \leftarrow x[j]$ // deplasare element $x[j]$ pe poziția $j + 1$
 2. $j \leftarrow j - 1$
 - endwhile**
 4. $x[j + 1] \leftarrow \text{aux}$ // deplasare element $x[i]$ pe poziția căutată
- endfor**
2. **return** x

Sortare prin inserție - Eficiență

Sort-Inserție(array x)

1. **for** $i = 2 \cdots n$
 1. $\text{aux} \leftarrow x[i]$
 2. $j \leftarrow i - 1$
 3. **while** $(j \geq 1)$ and $(\text{aux} < x[j])$
 1. $x[j + 1] \leftarrow x[j]$
 2. $j \leftarrow j - 1$**endwhile**
 4. $x[j + 1] \leftarrow \text{aux}$**endfor**
2. **return** x

Operații:

- **Comparații** $T_C(n)$
- **Deplasări elemente** $T_D(n)$

Pentru fiecare i :

$$1 \leq T_i(n) \leq i$$

Pentru toate valorile i :

$$n - 1 \leq T_C(n) \leq \frac{n(n + 1)}{2} - 1$$

Sortare prin inserție - Eficiență

Sort-Inserție(array x)

1. **for** $i = 2 \cdots n$
 1. $\text{aux} \leftarrow x[i]$
 2. $j \leftarrow i - 1$
 3. **while** $(j \geq 1)$ and $(\text{aux} < x[j])$
 1. $x[j + 1] \leftarrow x[j]$
 2. $j \leftarrow j - 1$**endwhile**
 4. $x[j + 1] \leftarrow \text{aux}$**endfor**
2. **return** x

Operații:

- **Comparații** $T_C(n)$
- **Deplasări elemente** $T_D(n)$

Pentru fiecare i :

$$0 + 2 \leq T_i(n) \leq (i - 1) + 2 = i + 1$$

Pentru toate valorile i :

$$2(n - 1) \leq T_D(n) \leq \frac{(n + 1)(n + 2)}{2} - 3$$

Sortare prin inserție - Eficiență

- Comparații: $n - 1 \leq T_C(n) \leq \frac{n(n+1)}{2} - 1$

- Deplasări elemente: $2(n - 1) \leq T_D(n) \leq \frac{(n+1)(n+2)}{2} - 3$

- Complexitate totală:

$$3(n - 1) \leq T(n) = T_C(n) + T_D(n) \leq n^2 + 2n - 3 \approx \mathcal{O}(n^2)$$

- Cel mai favorabil caz: $1, 2, 3, \dots, n$; Cel mai nefavorabil: $n, n - 1, \dots, 1$
- Algoritmul de sortare prin inserție este **stabil**!
- Este foarte simplu de implementat (există implementări în 5 linii de cod)!
- Recomandat pentru șiruri restrânse de elemente (performanțe practice bune!)

Sortare prin inserție binară

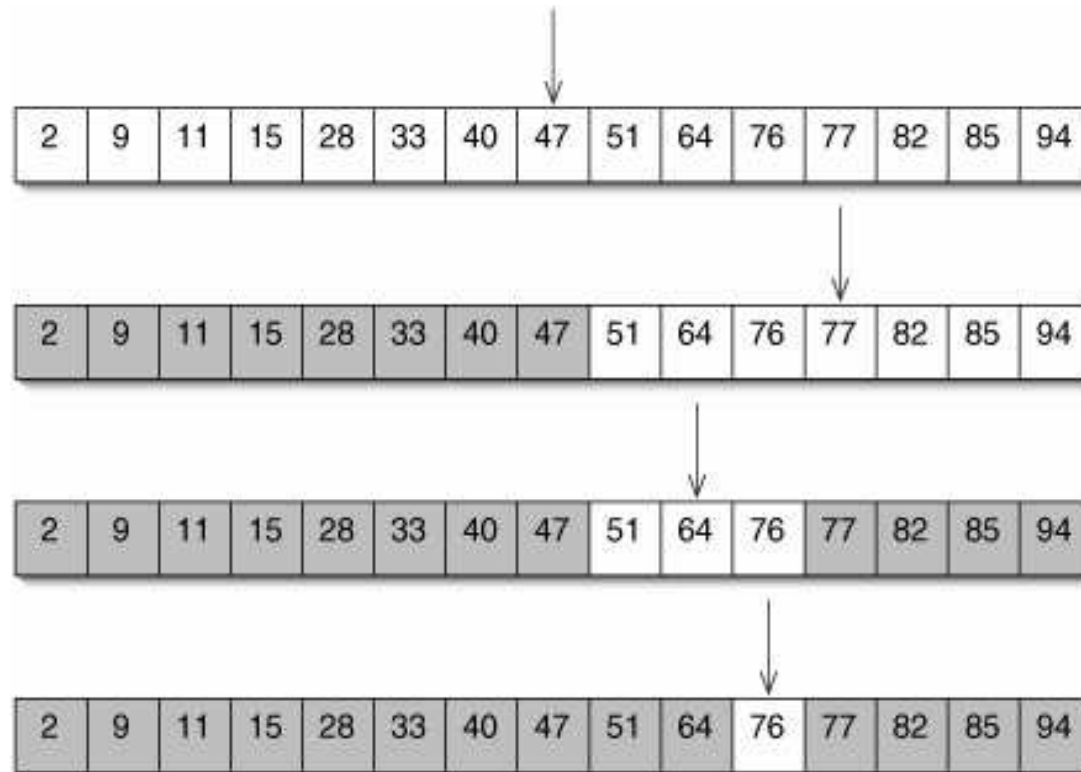
Idei generale:

- Considerăm șirul: $x[1], x[2], \dots, x[n]$.
- Se parcurge secvențial șirul (element cu element). Se inserează elementul curent ($x[i]$) în subșirul care îl precede ($x[1], x[2], \dots, x[i-1]$), folosind algoritmul de căutare binară, astfel încât acesta să rămână ordonat:

$$x[1], x[2], \dots, x[i-1], x[i], x[i+1], \dots, x[n]$$

- Îmbunătățire a metodei de sortare prin inserție directă: se înlocuiește procedura de căutare liniară ($\mathcal{O}(n)$) cu cea de căutare binară ($\mathcal{O}(\log(n))$).

Sortare prin inserție binară



<http://flylib.com/books/en/2.300.1.75/1/>

Sortare prin inserție binară

Căutare-binară(array x , int p , int q , int key):

1. $\text{mid} = p + ((q-p) / 2)$
2. **if** (key > $x[\text{mid}]$)
 1. return **Căutare-binară**(x , $\text{mid}+1$, q , key)**else**
 1. return **Căutare-binară**(x , p , mid , key)**endif**
3. **return** mid

- Complexitate: $T(n) = \mathcal{O}(\log(n))$
- Sortarea prin inserție binară reduce numărul de comparații de la $\mathcal{O}(n^2)$ la $\mathcal{O}(n \log(n))$
- Cu toate astea, complexitatea totală se păstrează $\mathcal{O}(n^2)$, datorită numărului de deplasări!

Cuprins

- Problema sortării
- Algoritmul de sortare prin interschimbare (Bubble sort)
- Algoritmul de sortare prin inserție (Insertion sort)
- **Algoritmul de sortare prin selecție (Selection sort)**
- Algoritmul de sortare prin interclasare (Mergesort)
- Algoritmul de sortare rapidă (Quicksort)

Sortare prin selecție (Selection sort)

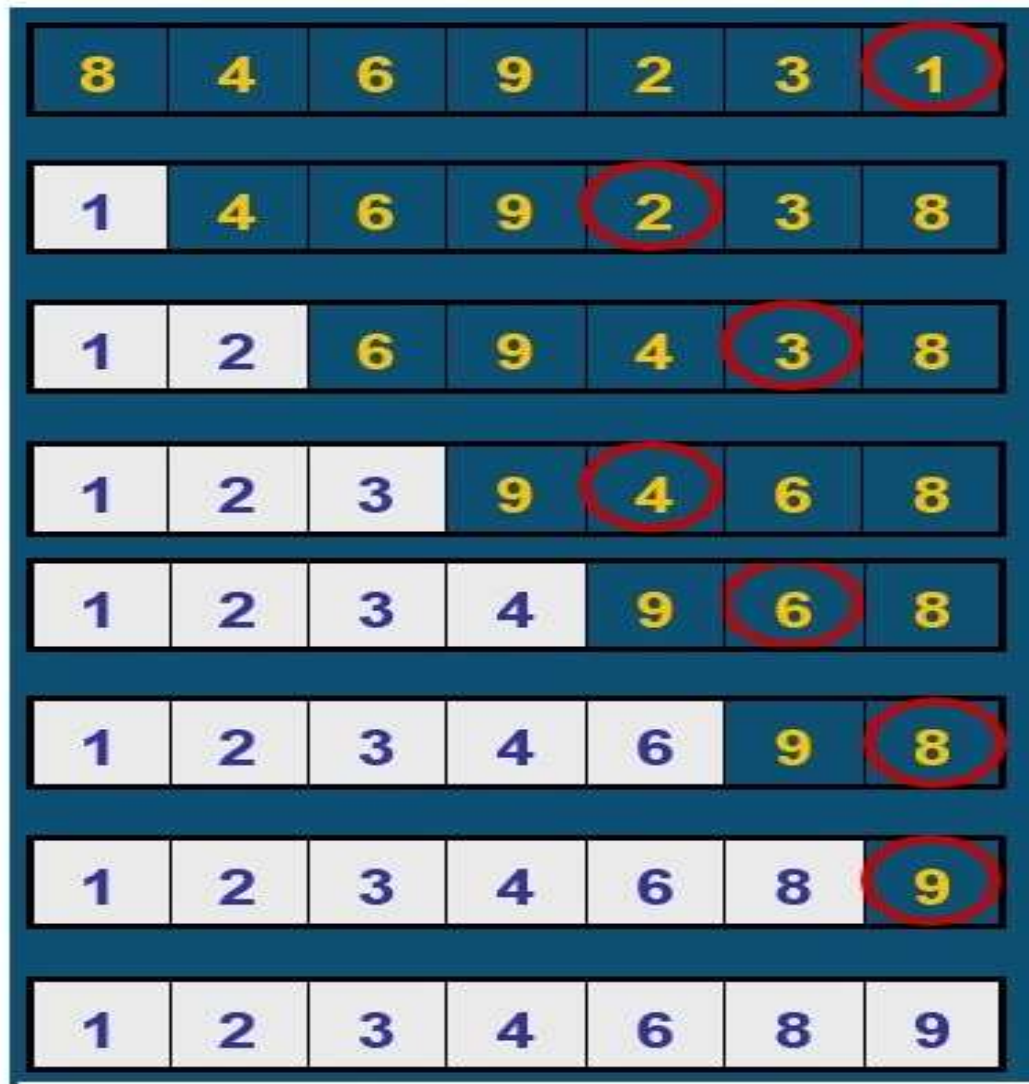
Idei generale:

- Considerăm șirul: $x[1], x[2], \dots, x[n]$.
- Se parcurge secvențial șirul (element cu element). Se determină minimum-ul din subșirul $(x[i], x[i + 1], \dots, x[n])$, și se înlocuiește cu elementul curent ($x[i]$).

$$[x[1], x[2], \dots, x[i - 1], x[i], x[i + 1], \dots, x[n]]$$

- Subșirul ce conține elementele deja sortate se mărește la fiecare iterație, astfel încât, după ce parcurgem toate elementele, secvența este sortată în întregime

Sortare prin selecție - Exemplu



Sortare prin selecție

Algoritm **Sort-Selecție**(array x):

1. **for** $i = 1 \cdots n - 1$
 1. $k \leftarrow i$
 2. **for** $j = i + 1 \cdots n$ // căutare minim în subșirul $x[i : n]$
 1. **if** $x[k] > x[j]$
 1. $k \leftarrow j$
 - endfor**
 3. **if** $k \neq i$ // dacă elementul curent nu este minim-ul subșirului $x[i : n]$
 1. $x[k] \leftrightarrow x[i]$ // interschimbare minim cu elementul curent
 - endfor**
2. **return** x

Sortare prin selecție - Eficiență

Sort-Selecție(array x):

1. **for** $i = 1 \cdots n - 1$
 1. $k \leftarrow i$
 2. **for** $j = i + 1 \cdots n$
 1. **if** $x[k] > x[j]$
 1. $k \leftarrow j$
 - endfor**
 3. **if** $k \neq i$
 1. $x[k] \leftrightarrow x[i]$
- endfor**
2. **return** x

Operații:

- **Comparații** $T_C(n)$
- **Deplasări elemente** $T_D(n)$

Pentru fiecare i :

$$T_i(n) = n - i$$

Pentru toate valorile i :

$$T_C(n) = \frac{n(n-1)}{2}$$

Sortare prin selecție - Eficiență

Sort-Selecție(array x):

1. **for** $i = 1 \cdots n - 1$
 1. $k \leftarrow i$
 2. **for** $j = i + 1 \cdots n$
 1. **if** $x[k] > x[j]$
 1. $k \leftarrow j$
 - endfor**
 3. **if** $k \neq i$
 1. $x[k] \leftrightarrow x[i]$
- endfor**
2. **return** x

Operații:

- **Comparații** $T_C(n)$
- **Deplasări elemente** $T_D(n)$

Pentru fiecare i :

$$0 \leq T_i(n) \leq 3$$

- fiecare interschimbare necesită 3 deplasări

Pentru toate valorile i :

$$0 \leq T_D(n) \leq 3(n - 1)$$

Algoritmul de sortare prin selecție - Eficiență

- Comparații: $T_C(n) = \frac{n(n-1)}{2}$
- Deplasări elemente: $0 \leq T_D(n) \leq 3(n-1)$
- Complexitate totală:

$$\frac{n(n-1)}{2} \leq T(n) = T_C(n) + T_D(n) \leq \frac{n(n-1)}{2} + 3(n-1) \approx \mathcal{O}(n^2)$$

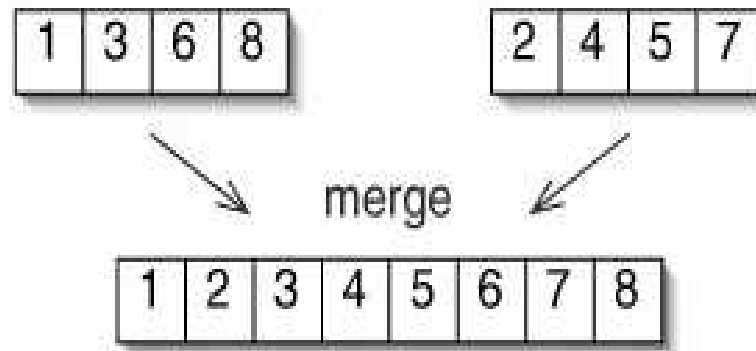
- Număr redus de deplasări elemente! (față de Sortarea prin inserție)
- Algoritmul de sortare prin selecție **nu este stabil!**

Cuprins

- Problema sortării
- Algoritmul de sortare prin interschimbare (Bubble sort)
- Algoritmul de sortare prin inserție (Insertion sort)
- Algoritmul de sortare prin selecție (Selection sort)
- **Algoritmul de sortare prin interclasare (Mergesort)**
- Algoritmul de sortare rapidă (Quicksort)

Interclasare

- Fie șirurile: $x[1], x[2], \dots, x[n]$ și $y[1], y[2], \dots, y[m]$ ($m \approx n$) ordonate crescător;
- Procedura de interclasare
 1. Se compară elementele minime din fiecare șir
 2. Se returnează cel mai mic
 3. Se repetă pașii precedenți până la epuizarea elementelor



Interclasare

Algorithm **Interclasare**(array x , array y)

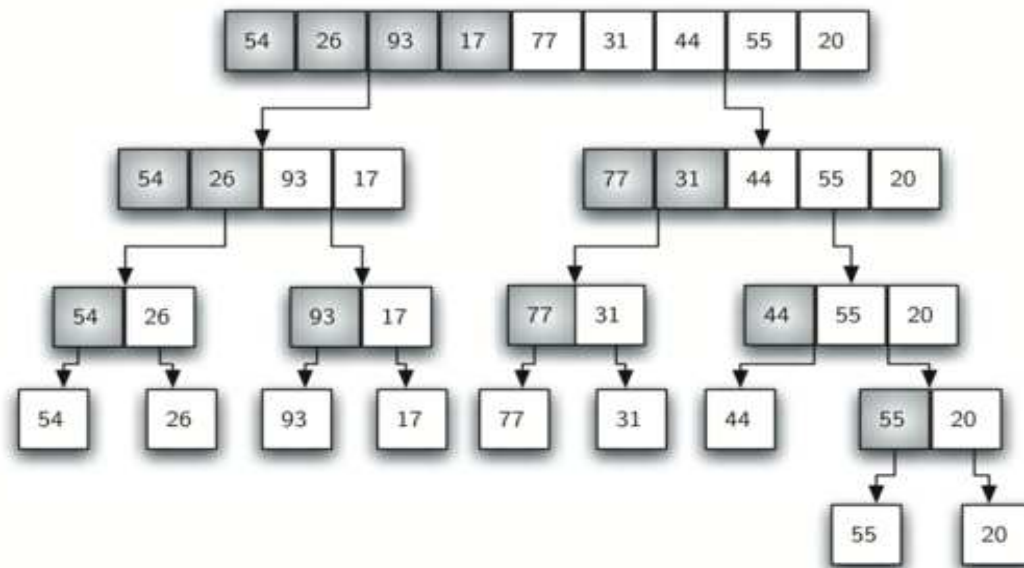
1. $i \leftarrow 1, j \leftarrow 1, k \leftarrow 1$
2. **while** $j \leq m$ and $i \leq n$
 2. **if** $x_i \leq y_j$
 1. $z_k \leftarrow x_i$
 2. $k \leftarrow k + 1$
 3. $i \leftarrow i + 1$
 - else**
 1. $z_k \leftarrow y_j$
 2. $k \leftarrow k + 1$
 3. $j \leftarrow j + 1$
3. **if** $i < n$
 1. $(z_k, \dots, z_{m+n}) \leftarrow (x_i, \dots, x_n)$
- else**
 1. $(z_k, \dots, z_{m+n}) \leftarrow (y_j, \dots, y_m)$
4. **return** z

- Complexitate $\mathcal{O}(m + n)$
- Operație mult mai simplă decât sortarea
- Reducem procesul de sortare la operații de interclasare: se interclasează subșiruri din ce în ce mai lungi, până se obține șirul sortat.

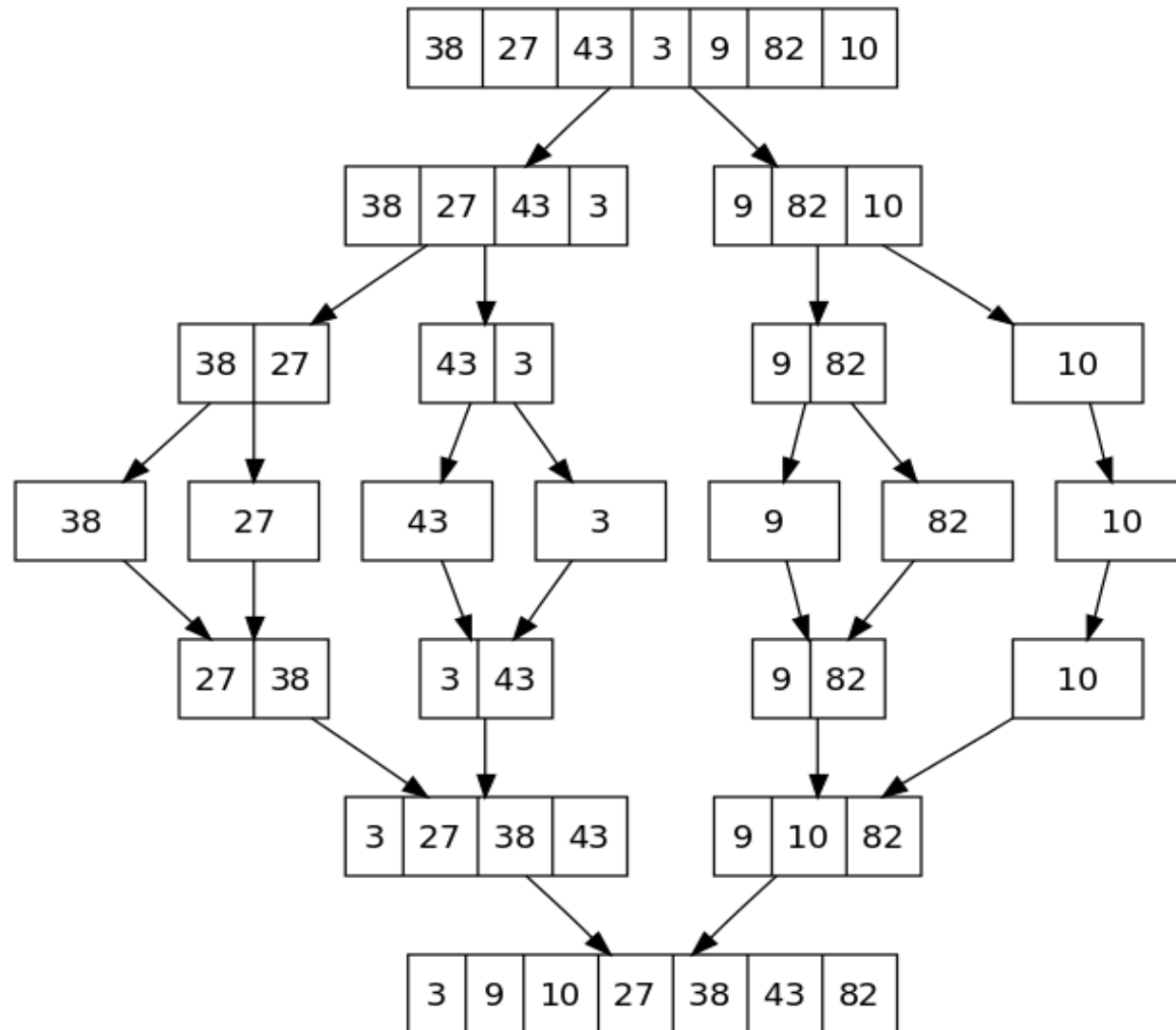
Sortare prin interclasare (Merge sort)

Structură generală:

- Considerăm șirul: $x[1], x[2], \dots, x[n]$.
- Se descompune șirul inițial nesortat în n subșiruri de lungime 1
- Se interclasează subșirurile obținute la pasul precedent până rezultă un singur șir



Sortare prin interclasare - Exemplu



Algoritmul de sortare prin interclasare

Algoritm **Sort-Interclasare**(array x , int p , int q):

1. **if** ($p < r$) // we have at least 2 items
 1. $q = (p + r) / 2$
 2. **Sort-Interclasare**(x , p , q) // sortare $x[p:q]$
 3. **Sort-Interclasare**(x , $q+1$, r) // sortare $x[q+1:r]$
 4. **Interclasare**($x[p:q]$, $x[q+1:r]$) // interclasare subșirurilor
- endif**

Sortare prin interclasare - Eficiență

Algoritm **Interclasare**(array x , array y)

1. $i \leftarrow 1, j \leftarrow 1, k \leftarrow 1$
2. **while** $j \leq m$ and $i \leq n$
 2. **if** $x_i \leq y_j$
 1. $z_k \leftarrow x_i$
 2. $k \leftarrow k + 1$
 3. $i \leftarrow i + 1$
 - else**
 1. $z_k \leftarrow y_j$
 2. $k \leftarrow k + 1$
 3. $j \leftarrow j + 1$
4. **if** $i > n$
 1. $(z_k, \dots, z_{m+n}) \leftarrow (x_i, \dots, x_n)$
- else**
 1. $(z_k, \dots, z_{m+n}) \leftarrow (y_j, \dots, y_m)$

- Operațiile dominante ale algoritmului **Sort-Interclasare** au loc în procedura de interclasare

- Se poate estima iterativ complexitatea $T(n)$:

- Problema este descompusă iterativ în două subprobleme cu dimensiune redusă la jumătate
- O subproblemă de dimensiune 1 costă $\mathcal{O}(1)$ operații
- Procedura de interclasare costă $\mathcal{O}(n)$

Sortare prin interclasare - Eficiență

- Complexitate totală: $\mathcal{O}(n \log(n))$

$$T(n) = \begin{cases} 1, & \text{dacă } n = 1 \\ T(\lceil \frac{n}{2} \rceil) + T(\lceil \frac{n}{2} \rceil) + n, & \text{altfel.} \end{cases}$$

- Recomandat pentru șiruri de mari dimensiuni
- Cu toate că are o complexitate mai bună, în cazul șirurilor de dimensiune mici, metoda sortării prin inserție prezintă performanțe mai bune decât sortarea prin interclasare
- Algoritmul de sortare prin interclasare **este stabil!**

Cuprins

- Problema sortării
- Algoritmul de sortare prin interschimbare (Bubble sort)
- Algoritmul de sortare prin inserție (Insertion sort)
- Algoritmul de sortare prin selecție (Selection sort)
- Algoritmul de sortare prin interclasare (Mergesort)
- **Algoritmul de sortare rapidă (Quicksort)**

Algoritmul de sortare rapidă (Quicksort)

- Algoritm de sortare eficient dezvoltat de Tony Hoare în 1959 și publicat în 1961
- Implementările bune pot depăși de 2-3 ori performanțele sortării prin interclasare

Idei generale:

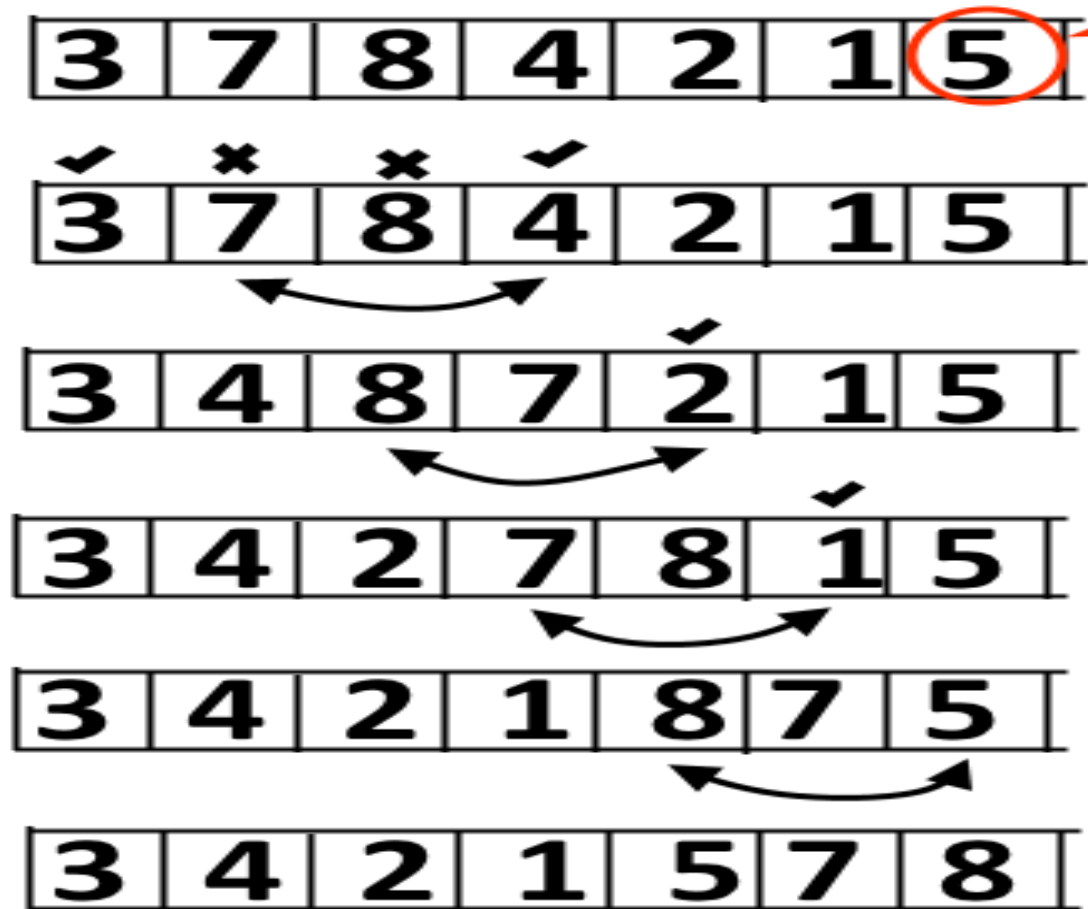
- Considerăm șirul: $x[1], x[2], \dots, x[n]$.
- Se alege din șir un element, numit *pivot*.

$$[x[1], x[2], \dots, x[i-1], x[i], x[i+1], \dots, x[n]]$$

Se realizează *partiționarea* șirului (în raport cu acest pivot): se reordonează șirul astfel încât **elementele cu valori mai mici** decât pivotul se plasează înainte de pivot, iar **elementele cu valori mai mari** decât pivotul se plasează după acesta (cele egale în oricare din părți).

- După partiționare, pivotul se află în poziția finală. Se aplică recursiv aceeași procedură subșirului de elemente cu valori mai mici, iar separat subșirului cu valori mai mari.

Algoritmul de sortare rapidă (Quicksort)



Algoritmul de sortare rapidă (Quicksort)

Algoritm **Quicksort** (array x , int l , int h)

1. if ($l < h$)
 1. LocPivot = **Partitionare**(x, l, h)
 2. **Quicksort**(x, l , LocPivot) // recursie pe segmentul elementelor mai mici
 3. **Quicksort**(x , LocPivot + 1, h) // recursie segment elemente mai mari

Algoritm **Partitionare** (array x , int l , int h)

1. pivot = $x[l]$ // considerăm pivot primul element
2. leftwall = l
3. **for** $i = l + 1 \dots h$
 - (a) **if** ($x[i] < \text{pivot}$)
 1. $x[i] \leftrightarrow x[\text{leftwall}]$ // elementul $x[i]$ se mută în segmentul corect
 2. leftwall = leftwall + 1
4. pivot $\leftrightarrow x[\text{leftwall}]$
5. **return** leftwall

Algoritmul de sortare rapidă (Quicksort)

Algoritm **Quicksort** (array x , int l , int h)

1. if ($l < h$)
 1. LocPivot = **Partitionare**(x, l, h)
 2. **Quicksort**(x, l , LocPivot)
 3. **Quicksort**(x , LocPivot + 1, h)

Algoritm **Partitionare** (array x , int l , int h)

1. pivot = $x[l]$
2. leftwall = l
3. **for** $i = l + 1 \dots h$
 - (a) **if** ($x[i] < \text{pivot}$)
 1. $x[i] \leftrightarrow x[\text{leftwall}]$
 2. leftwall = leftwall + 1
4. pivot $\leftrightarrow x[\text{leftwall}]$
5. **return** leftwall

- Operațiile dominante ale algoritmului **Quicksort** au loc în procedura de partiționare

- Estimare complexitate $T(n)$:

- Problema este descompusă iterativ în două subprobleme cu dimensiune redusă la jumătate
- O subproblemă de dimensiune 1 costă $\mathcal{O}(1)$ operații
- Procedura de partiționare costă, în cel mai rău caz, $\mathcal{O}(n)$ comparații

Sortare prin interclasare - Eficiență

- Complexitate în cel mai rău scenariu: $\mathcal{O}(n^2)$
- Complexitate în medie: $\mathcal{O}(n \log(n))$

Avantaje:

- Necesită memorie $\mathcal{O}(\log(n))$; mult mai redusă decât sortarea prin interclasare
- Reprezintă cea mai bună alegere când viteza este foarte importantă, indiferent de dimensiunea setului de date.

Dezavantaje:

- Poate duce la umplerea stivei când se utilizează seturi largi de date.
- Performanțe modeste atunci când operează asupra unor liste aproape ordonate.
- Algoritmul de sortare rapidă **este instabil!**

Concluzii

- O mulțime de metode de sortare bazate pe comparații între chei, cu avantaje și dezavantaje proprii
- Aplicația practică dictează alegerea metodei
- Metodele de sortare prin inserție și selecție au performanțe bune când șirul de sortat are dimensiuni reduse
- Pentru șiruri de dimensiuni mari: sortare prin interclasare (Mergesort), sortare rapidă (Quicksort), etc.