

---

**Algoritmi și structuri de date (I). Seminar 7:** Variante ale algoritmilor de sortare: sortare prin numărare, sortare pe baza cifrelor, shell sort, shaker sort.

---

**Problema 1** *Counting sort - sortare prin numărare.* Să considerăm problema sortării unui tablou  $x[1..n]$  având elemente din  $\{1, 2, \dots, m\}$ . Pentru acest caz particular poate fi utilizat un algoritm de complexitate liniară bazat pe următoarele idei:

- se construiește tabelul  $f[1..m]$  al frecvențelor de apariție a valorilor elementelor tabloului  $x[1..n]$ ;
- se calculează frecvențele cumulate asociate;
- se folosește tabelul frecvențelor cumulate pentru a construi tabloul ordonat  $y$ .

Algoritmul poate fi descris prin:

---

```
countingsort(integer  $x[1..n]$ ,  $m$ )
integer  $i$ ,  $f[1..m]$ ,  $y[1..n]$ 
for  $i \leftarrow 1, m$  do  $f[i] \leftarrow 0$  endfor
for  $i \leftarrow 1, n$  do  $f[x[i]] \leftarrow f[x[i]] + 1$  endfor
for  $i \leftarrow 2, m$  do  $f[i] \leftarrow f[i - 1] + f[i]$  endfor
for  $i \leftarrow n, 1, -1$  do
     $y[f[x[i]]] \leftarrow x[i]$ 
     $f[x[i]] \leftarrow f[x[i]] - 1$ 
endfor
for  $i \leftarrow 1, n$  do  $x[i] \leftarrow y[i]$  endfor
return  $x[1..n]$ 
```

---

Stabiliți ordinul de complexitate al algoritmului `countingsort`.

*Indicație.* Dimensiunea problemei este determinată de perechea  $(m, n)$ .

**Problema 2** *Radix sort - sortare pe baza cifrelor.* Considerăm un tablou  $x[1..n]$  având elemente numere naturale constituite din cel mult  $k$  cifre. Dacă  $m = 10^k$  este semnificativ mai mare decât  $n$  atunci sortarea prin numărare nu este eficientă. Pe de altă parte, dacă  $k$  este mult mai mic decât  $n$  atunci un algoritm de complexitate  $\mathcal{O}(kn)$  ar putea fi acceptabil. Un astfel de algoritm este cel bazat pe următoarea idee: se ordonează tabloul în raport cu cifra cea mai puțin semnificativă a fiecărui număr (folosind counting sort în ipoteza că mulțimea valorilor din tabloul de sortat este mulțimea cifrelor,  $\{0, 1, \dots, 9\}$ ) după care se sortează în raport cu cifra de rang imediat superior ș.a.m.d. până se ajunge la cifra cea mai semnificativă. Deci structura generală a algoritmului este:

---

```
radixsort(integer  $x[1..n]$ ,  $k$ )
integer  $i$ 
for  $i \leftarrow 0, k - 1$  do
     $x[1..n] \leftarrow \text{counting2}(x[1..n], 9, i)$ 
endfor
return  $x[1..n]$ 
```

---

Algoritmul `counting2` este o adaptare a algoritmului de sortare prin numărare în raport cu cifrele de un anumit rang ( $i$ ):

---

```

counting2(integer  $x[1..n], m, c$  )
integer  $i, j, f[0..m], y[1..n]$ 
for  $i \leftarrow 0, m$  do  $f[i] \leftarrow 0$  endfor
for  $i \leftarrow 1, n$  do
     $j \leftarrow (x[i] \text{DIV } \text{putere}(10, c)) \text{MOD } 10$ 
     $f[j] \leftarrow f[j] + 1$ 
endfor
for  $i \leftarrow 1, m$  do  $f[i] \leftarrow f[i-1] + f[i]$  endfor
for  $i \leftarrow n, 1, -1$  do
     $j \leftarrow (x[i] \text{DIV } \text{putere}(10, c)) \text{MOD } 10$ 
     $y[f[j]] \leftarrow x[i]$ 
     $f[j] \leftarrow f[j] - 1$ 
endfor
for  $i \leftarrow 1, n$  do  $x[i] \leftarrow y[i]$  endfor
return  $x[1..n]$ 

```

---

Stabiliți ordinul de complexitate al algoritmului **radixsort**.

**Problema 3** (*Shell sorting - sortare prin inserție cu pas variabil.*) Unul dintre dezavantajele sortării prin inserție este faptul că la fiecare etapă un element al șirului se deplasează cu o singură poziție. O variantă de reducere a numărului de operații efectuate este de a compara elemente aflate la o distanță mai mare ( $h \geq 1$ ) și de a realiza deplasarea acestor elemente peste mai multe poziții. De fapt tehnica se aplică în mod repetat pentru valori din ce în ce mai mici ale pasului  $h$ , asigurând  $h$ -sortarea șirului. Un șir  $x[1..n]$  este considerat  $h$ -sortat dacă orice subșir  $x[i_0], x[i_0 + h], x[i_0 + 2h] \dots$  este sortat ( $i_0 \in \{1, \dots, h\}$ ). Aceasta este ideea algoritmului propus de Donald Shell în 1959 cunoscut sub numele "shell sort".

Elementul cheie al algoritmului îl reprezintă alegerea valorilor pasului  $h$ . Pentru alegeri adecvate ale secvenței  $h_k$  se poate obține un algoritm de complexitate  $\mathcal{O}(n^{3/2})$  în loc de  $\mathcal{O}(n^2)$  cum este în cazul algoritmului clasic de sortare prin inserție.

*Exemplu.* Exemplificăm ideea algoritmului în cazul unui șir cu 15 elemente pentru următoarele valori ale pasului:  $h = 13, h = 4, h = 1$  (care corespund unui șir  $h_k$  dat prin relația  $h_k = 3h_{k-1} + 1, h_1 = 1$ ):

*Etapă 1:* pentru  $h = 13$  se aplică algoritmul sortării prin inserție subșirurilor  $x[1], x[14]$  și  $x[2], x[15]$ , singurele subșiruri cu elemente aflate la distanța  $h$  care au mai mult de un element:

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
<span style="border: 1px solid black; padding: 0 2px;">14</span>	8	10	7	9	12	5	15	2	13	6	1	4	<span style="border: 1px solid black; padding: 0 2px;">3</span>	11

și se obține

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
3	8	10	7	9	12	5	15	2	13	6	1	4	14	11

*Etapă 2:* pentru  $h = 4$  se aplică algoritmul sortării prin inserție succesiv subșirurilor:  $x[1], x[5], x[9], x[13], x[2], x[6], x[10], x[14], x[3], x[7], x[11], x[15], x[4], x[8], x[12]$ . După prima subetapă (prelucrarea primului subșir) prin care se ordonează subșirul constituit din elementele marcate:

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
<span style="border: 1px solid black; padding: 0 2px;">3</span>	8	10	7	<span style="border: 1px solid black; padding: 0 2px;">9</span>	12	5	15	<span style="border: 1px solid black; padding: 0 2px;">2</span>	13	6	1	<span style="border: 1px solid black; padding: 0 2px;">4</span>	14	11

se obține:

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
2	8	10	7	3	12	5	15	4	13	6	1	9	14	11

La a doua subetapă se aplică sortarea prin inserție asupra subșirului constituit din elementele marcate:

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
2	8	10	7	3	12	5	15	4	13	6	1	9	14	11

obținându-se aceeași configurație (subșirul este deja ordonat crescător) din care se prelucrează acum subșirul constituit din elementele marcate:

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
2	8	10	7	3	12	5	15	4	13	6	1	9	14	11

obținându-se

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
2	8	5	7	3	12	6	15	4	13	10	1	9	14	11

Se aplică acum sortarea prin inserție asupra subșirului constituit din elementele marcate:

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
2	8	5	7	3	12	6	15	4	13	10	1	9	14	11

obținându-se

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
2	8	5	1	3	12	6	7	4	13	10	15	9	14	11

Se aplică acum sortarea prin inserție asupra întregului șir. Pentru acest exemplu aplicarea directă a sortării prin inserție conduce la execuția a 68 de comparații pe când aplicarea sortării cu pas variabil de inserție conduce la execuția a 34 de comparații.

Prelucrarea specifică unei etape (o valoare a pasului de inserție) poate fi descrisă prin:

---

```

insertie_pas(real  $x[1..n]$ , integer  $h$ )
integer  $i, j$ 
real  $aux$ 
for  $i \leftarrow h + 1, n$  do
     $aux \leftarrow x[i]$ 
     $j \leftarrow i - h$ 
    while  $j \geq 1$  AND  $aux < x[j]$  do
         $x[j + h] \leftarrow x[j]$ 
         $j \leftarrow j - h$ 
    endwhile
     $x[j + h] \leftarrow aux$ 
endfor
return  $x[1..n]$ 

```

---

Această prelucrare se efectuează pentru diferite valori ale lui  $h$ . Un exemplu de șir de valori pentru  $h$  (care s-a dovedit a conduce la o variantă eficientă a algoritmului) este cel bazat pe relația de recurență:  $h_k = 3h_{k-1} + 1$ ,  $h_1 = 1$ . Structura generală a algoritmului va consta în:

---

```

shellsort(real  $x[1..n]$ )
integer  $h$ 
 $h \leftarrow 1$ 
while  $h \leq n$  do
     $h \leftarrow 3 * h + 1$ 
endwhile
repeat
     $h \leftarrow h \text{DIV} 3$ 
     $x[1..n] \leftarrow \text{insertie\_pas}(x[1..n], h)$ 
until  $h = 1$ 
return  $x[1..n]$ 

```

---

O altă variantă de alegere a pasului  $h$ , pentru care s-a demonstrat că algoritmul are complexitatea  $\mathcal{O}(n\sqrt{n})$  este  $h_k = 2^k - 1$ .

**Problema 4** (*Shaker sort.*) Sortarea prin interschimbarea elementelor vecine asigură, la fiecare pas al ciclului exterior, plasarea câte unui element (de exemplu maximul din subtabloul tratat la etapa respectivă) pe poziția finală. O variantă ceva mai eficientă ar fi ca la fiecare etapă să se plaseze pe pozițiile finale câte două elemente (minimul respectiv maximul din subtabloul tratat la etapa respectivă). Pe de altă parte prin reținerea indicelui ultimei interschimbări efectuate, atât la parcurgerea de la stânga la dreapta cât și de la dreapta la stânga se poate limita regiunea analizată cu mai mult de o poziție atât la stânga cât și la dreapta (când tabloul conține porțiuni deja sortate).

Structura algoritmului este:

---

```

shakersort(real  $x[1..n]$ )
integer  $s, d, i, t$ 
 $s \leftarrow 1; d \leftarrow n$ 
repeat
     $t \leftarrow 0$ 
    for  $i \leftarrow s, d - 1$  do
        if  $x[i] > x[i + 1]$  then  $x[i] \leftrightarrow x[i + 1]; t \leftarrow i$  endif
    endfor
    if  $t \neq 0$  then
         $d \leftarrow t; t \leftarrow 0$ 
        for  $i \leftarrow d, s + 1, -1$  do
            if  $x[i] < x[i - 1]$  then  $x[i] \leftrightarrow x[i - 1]; t \leftarrow i$  endif
        endfor
         $s \leftarrow t$ 
    endif until  $t = 0$  OR  $s = d$ 
return  $x[1..n]$ 

```

---

## Probleme suplimentare

1. Propuneți un algoritm care sortează crescător elementele de pe pozițiile impare ale unui tablou și descrescător cele de pe poziții pare.
2. Adaptați algoritmul de sortare prin numărare pentru ordonarea descrescătoare a unui tablou.
3. Se consideră un set de  $n$  valori din  $\{1, \dots, m\}$ . Preprocesați acest set folosind un algoritm de complexitate  $\mathcal{O}(\max\{m, n\})$  astfel încât răspunsul la întrebarea "câte elemente se află în intervalul  $[a, b]$ ,  $a, b \in \{1, \dots, m\}$  să poată fi obținut în timp constant. *Indicație.* Se poate construi un tablou adițional.
4. Propuneți un algoritm de complexitate liniară pentru ordonarea crescătoare a unui tablou constituit din  $n$  valori întregi aparținând mulțimii  $\{0, 2, \dots, n^2 - 1\}$ . *Indicație.* Se vor interpreta valorile ca fiind reprezentate în baza  $n$ .
5. Se consideră un tablou constituit din triplete de trei valori întregi corespunzătoare unei date calendaristice (zi, luna, an). Propuneți un algoritm de ordonare crescătoare după valoarea datei.
6. Se consideră o succesiune de  $2n$  discuri de culori alternate: un disc roșu este urmat de un disc albastru, iar un disc albastru este urmat de un disc roșu (ARARAR... sau RARARA...). Se pune problema ca toate discurile de aceeași culoare să fie grupate (de exemplu AAARRR sau RRRAAA) în condițiile în care singura operație permisă este interschimbarea a două discuri vecine. Propuneți un algoritm de rezolvare a problemei și stabiliți ordinul de complexitate al algoritmului.
7. Se consideră un set de  $n$  puncte în plan specificate prin coordonatele lor carteziane  $((x_1, y_1), (x_2, y_2), \dots, (x_n, y_n))$ . Să se ordoneze crescător setul de puncte în funcție de distanța euclidiană față de originea sistemului de axe de coordonate. *Indicație.* Distanța euclidiană de la un punct de coordonate  $(x_i, y_i)$  la originea sistemului de coordonate este  $\sqrt{x_i^2 + y_i^2}$ .
8. Se consideră un set de  $n$  localități plasate de-a lungul unui drum drept, astfel că pozițiile localităților pot fi specificate ca fiind coordonatele  $0 \leq x_1 < x_2 < \dots < x_n$ . Se pune problema selectării unei localități în care să se plaseze o școală astfel încât: (i) să fie minimizată *distanța maximă* dintre localitatea unde e plasată școală și fiecare dintre celelalte localități; (ii) să fie minimizată *media distanțelor* dintre localitatea unde e plasată școală și fiecare dintre celelalte localități.