SDA – Seminar 5 - Ansamblu

- **Conținut**: Materialul curent exemplifică aplicabilitatea structurii de date **Ansamblu** (*Heap*) pentru rezolvarea unei probleme, prezentându-se și comparându-se din perspectiva eficienței timp soluții multiple.

Enunțul problemei

Determinați suma celor mai mari *k* elemente dintr-un vector conținând *n* elemente numerice distincte.

Exemplu

De exemplu, dacă vectorul conține următoarele 10 elemente: [6, 12, 91, 9, 3, 5, 25, 81, 11, 23] și k = 3, rezultatul așteptat este: 91 + 81 + 25 = 197.

Soluții posibile

I. Determinând maximul de k ori

Observații:

- Dacă, pentru exemplul considerat, aplicăm funcția de determinare a maximului de 3 ori, aceasta va returna 91 de fiecare dată. Așadar, este necesară impunerea unei limite superioare impuse maximului calculat, căutându-se astfel elementul maxim care este mai mic decât limita superioară impusă.
- Generalizând, începând cu cel de-al doilea pas, limita superioară impusă este maximul determinat la pasul anterior. La primul pas se poate alege ca limită o valoare mai mare decât marginea superioară a domeniului valorilor posibile.
- Pentru exemplul considerat:
  - Inițial, maximul este 91.
  - Prin cel de-al doilea apel al funcției de determinare a maximului, dorim determinarea maximului dintre elementele strict mai mici decât 91. Vom obține 81.
  - La cel de-al treilea apel, dorim maximul strict mai mic decât 81, obținând 25.

Care este complexitatea timp a acestei soluții?

- Complexitatea acestei soluții este: Θ(k*n) – găsirea maximului are loc în Θ(n) și se repetă de k ori.

II. Sortarea șirului în ordine descrescătoare și însumarea primelor k elemente.

Care este complexitatea timp a acestei soluții?

- Sortarea poate fi efectuată în $\Theta(n*\log_2 n)$ (folosind sortarea prin interclasare sau *HeapSort*) (OBS: dacă domeniul valorilor șirului ar respecta condițiile necesare aplicării *BucketSort,* s-ar putea efectua în timp liniar, însă neavând această asigurare, nu o vom presupune)
- Calculul sumei primelor k elemente din șirul ordonat descrescător în prealabil se efectuează în Θ(k)
- Obținem, astfel, complexitatea totală a soluției ca fiind $\Theta(n*\log_2 n) + \Theta(k) \in \Theta(n*\log_2 n)$

III. **Prin adăugarea tuturor celor n elemente într-un ansamblu binar maximal (*binary max-heap),* urmată de eliminarea și însumarea primelor k (celor mai mari) elemente.**

Care este complexitatea adăugării unui element într-un ansamblu cu m elemente?

- Complexitatea adăugării unui element într-un ansamblu cu m elemente este **$O(\log_2 m)$**.

Care este complexitatea ștergerii unui element dintr-un ansamblu cu m elemente?

- Complexitatea eliminării unui element dintr-un ansamblu cu m elemente este **$O(\log_2 m)$.**

Care este numărul maxim de elemente conținut de ansamblul în care adăugăm / din care ștergem elemente?

- Ansamblul în care adăugăm și din care ștergem elemente conține cel mult n elemente (anterior primei ștergeri conține exact n elemente).

Câte adăugări se efectuează? Dar ștergeri?

- Se efectuează n adăugări (fiecare dintre cele n elemente fiind adăugat în ansamblu) și k ștergeri (sau, eventual, k-1 ștergeri + accesarea maximului dintre elementele rămase în ansamblu și adăugarea lui sumei).

Care este complexitatea totală obținută?

- Obținem complexitatea totală: $O(n*\log_2 n) + O(k*\log_2 n)$. Având în vedere că n >= k, se obține $O(n*\log_2 n)$.

```
Funcție sumaCelorMaiMariK(sir, n, k):
//sir este un vector (tablou unidimensional) de numere întregi distincte
//n este dimensiunea vectorului sir (adică numărul de elemente pe care acesta le conține)
//k este numărul de elemente pe care dorim să le însumăm, verificându-se inegalitatea k <= n

      init(a, "≥") //presupunând că avem o structură de date Ansamblu implementată, a fiind o instanță a ei, inițializăm relația impusă ansamblului cu "≥", obținând un ansamblu maximal (max-heap)
      Pentru i ← 1, n execută //fiecare dintre cele n elemente din șir
                  adaugă(a, sir[i]) //este adăugat în ansamblu
      SfPentru
      suma ← 0 //inițializăm suma
      Pentru i ← 1, k execută //de k ori
```

```
                elem ← șterge(a) //eliminăm maximul curent din ansamblu
                suma ← suma + elem //și îl adăugăm sumei
        SfPentru
        sumaCelorMaiMariK ← suma //stabilim rezultatul algoritmului (funcției)
    SfFuncție
```

Putem reduce complexitatea soluției anterioare, $O(n*\log_2 n)$, printr-o soluție mai eficientă ?

IV. Observăm nu avem nevoie de toate cele n elemente în ansamblu, întrucât ne interesează doar cele mai mari k.

- Dacă reconsiderăm exemplul (șirul este [6, 12, 91, 9, 3, 5, 25, 81, 11, 23] și k = 3), **putem păstra în ansamblu, la fiecare pas, doar cele mai mari k elemente** întâlnite până la momentul curent, procedând după cum urmează:
- Inițial adăugăm primele k elemente în ansamblu. Deci, pentru exemplul considerat, în momentul în care ajungem cu parcurgerea șirului la elementul 9, ansamblul conține 6, 12 și 91.
- Când întâlnim 9, putem renunța la 6, știind sigur că nu va face parte din cele mai mari 3 numere, întrucât avem deja 3 numere mai mari decât acesta. Așadar, păstrăm 12, 91 și 9.
- Când întâlnim 3, știm că 3 nu va fi printre cele mai mari 3 numere, întrucât avem deja 3 elemente mai mari decât 3 în ansamblu. Conținutul ansamblului rămâne, în continuare: 12, 91 și 9. Similar pentru elementul 5.
- Când întâlnim 25, putem renunța la 9, continuând cu 12, 91 și 25.

Ce se întâmplă la următorii 3 pași și care vor fi cele 3 elemente conținute în final de ansamblu?

- Când întâlnim 81, putem renunța la 12, continuând cu 91, 25 și 81.
- Când întâlnim 11, continuăm cu 91, 25 și 81.
- Când întâlnim 23, păstrăm în ansamblu 91, 25 și 81.
- Astfel, ansamblul conține la finalul parcurgerii șirului elementele 91, 25 și 81, care sunt cele mai mari 3 elemente din întreg șirul.

La ce se reduce rezolvarea problemei ulterior parcurgerii întregului șir și actualizării corespunzătoare a conținutul ansamblului?

- Ulterior parcurgerii șirului și construirii ansamblului format din cele mai mari k elemente, rezolvarea problemei se reduce la însumarea elementelor componente ale ansamblului.

Ansamblul în care păstrăm cele mai mari k elemente din șir ar trebui să fie un ansamblu maximal (*max-heap*) sau unul minimal (*min-heap*)?

- Având în ansamblu, la un moment dat, cele mai mari k elemente anterioare elementului curent , suntem interesați de elementul minim dintre acestea, astfel încât să îl comparăm cu elementul curent. Dacă elementul curent este mai mare decât minimul elementelor din ansamblu, atunci îl va înlocui. Altfel, conținutul ansamblului rămâne nemodificat (știind că toate elementele din ansamblu sunt mai mari decât elementul curent). Prin urmare, avem nevoie de un **ansamblu minimal (*min-heap*).**

```
Functie sumaCelorMaiMariK_2(sir, n, k):
//sir este un vector (tablou unidimensional) de numere întregi distincte
//n este dimensiunea vectorului sir (adică numărul de elemente pe care acesta le conține)
//k este numărul de elemente pe care dorim să le însumăm, verificându-se inegalitatea k <= n

     init(a, "≤") //presupunând că avem o structură de date Ansamblu implementată, a fiind o instanță a ei, inițializăm relația impusă ansamblului cu "≤", obținând un ansamblu minimal(min-heap)
     Pentru i ← 1, k execută //primele k elemente ale șirului sunt adăugate în ansamblu în mod implicit
          adaugă(a, șir[i])
     SfPentru
     Pentru i ← k+1, n execută //fiecare element rămas în șir va fi comparat cu elementul minim din ansamblu
          Dacă sir[i] > prim(a) atunci //dacă este mai mare decât elementul minim din ansamblu, returnat de funcția prim, atunci
               șterge(a) //, elementul minim din ansamblu este eliminat
               adaugă(a, șir[i]) //, iar elementul curent este adăugat în ansamblu
```

```
            SfDacă
        //calculăm suma celor k elemente conținute de ansamblu la finalul parcurgerii șirului
    SfPentru
    Pentru i ← 1, k execută //de k ori
                    elem ← șterge(a) //eliminăm minimul curent din ansamblu
                    suma ← suma + elem //și îl adăugăm sumei
    SfPentru
    sumaCelorMaiMariK_2 ← suma
SfSFuncție
```

 Care este complexitatea timp a acestei soluții?

 Care este numărul maxim de elemente conținut de ansamblul în care adăugăm / din care ștergem elemente?

- Ansamblul în care adăugăm și din care ștergem elemente conține cel mult **k** elemente.

 Câte sunt complexitățile operațiilor de adăugare și ștergere ?

- Adăugările și ștergerile se efectuează în **$O(\log_2 k)$**.

 Câte adăugări și câte ștergeri se efectuează?

- Adăugăm și ștergem de cel mult **n** ori (de pildă, dacă șirul este ordonat crescător).
  - Toate cele n elemente ale șirului vor fi adăugate în ansamblu
  - Primele n-k elemente vor fi eliminate din ansamblu până la construirea ansamblului final, odată cu parcurgerea șirului, restul de k elemente fiind eliminate pentru calculul sumei

Care este complexitatea totală obținută?

- Obținem complexitatea totală: **O(n*log$_2$k)**.

Observații:

Dacă am lucra direct pe reprezentare, în ultima structură repetitivă *Pentru* în care eliminăm elemente din ansamblu și le însumăm, am putea doar să însumăm elementele din ansamblu. În acest caz, clasa de complexitate nu se schimbă, dar numărul efectiv de operații elementare efectuate se reduce.

V. Ne amintim (din Cursul 7) că putem să **transformăm șirul într-un ansamblu** în O(n). Prin urmare, putem proceda similar soluției III, dar în loc să adăugăm elementele unul câte unul în ansamblu, putem să transformăm șirul într-un ansamblu maximal **și apoi să eliminăm k elemente din el**.

Care este complexitatea acestei soluții?

- Complexitatea acestei soluții este: **O(n + k*log$_2$n)**.