SDA (PC2)
Curs 9
Arbori

Iulian Năstac

Lista dublu înlănțuită
Recapitulare

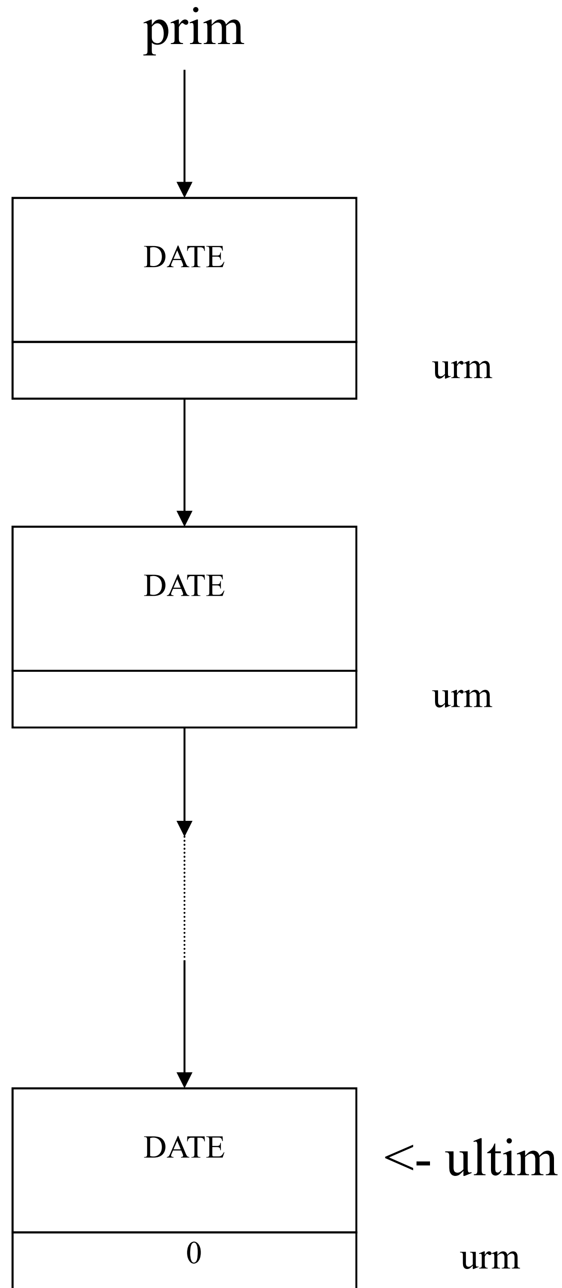
- Într-o astfel de listă fiecare nod conţine doi pointeri: unul spre nodul următor şi unul spre nodul precedent.

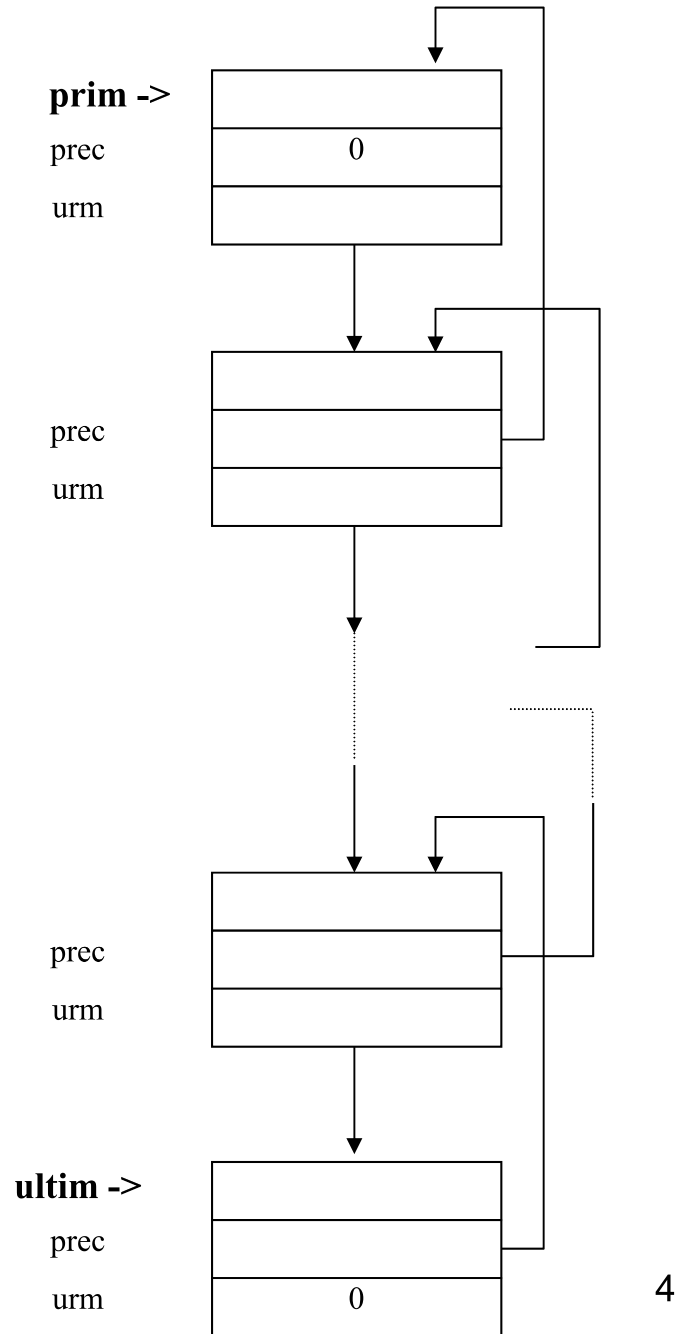
- Astfel fiecare nod al listei are un nod precedent definit prin pointerul **prec** şi un nod urmator definit prin pointerul **urm**.

Observaţii:

- Se consideră tipul utilizator:

```
typedef struct nod
        { declarații;
          struct nod *prec;
          struct nod *urm;
        } NOD;
```


prim
DATE
urm
DATE
urm
DATE
<- ultim
0
urm


prim ->
prec
0
urm
prec
urm
prec
urm
ultim ->
prec
urm
0

Grafuri
Recapitulare

Definiție:

Un graf este o pereche

$$G = <V, M>$$

unde V este o mulțime de vârfuri, iar $M \subseteq V \times V$ este o mulțime de muchii (laturi).

De obicei muchia de la vârful **a** la vârful **b** este notată cu:
- perechea ordonată **(a, b)** dacă graful este orientat;
- perechea neordonată **{a, b}** dacă graful este neorientat.

Observație: S-a presupus că ***a*** și ***b*** sunt diferite.

Un graf poate fi:

- orientat,
- neorientat,
- mixt.

Un **drum** este o succesiune de muchii de forma:

- $(a_1, a_2), (a_2, a_3), (a_3, a_4), \ldots , (a_{n-1}, a_n)$ dacă graful este orientat

- $\{a_1, a_2\}, \{a_2, a_3\}, \{a_3, a_4\}, \ldots , \{a_{n-1}, a_n\}$ dacă graful este neorientat

Definiții

- Lungimea unui drum = numărul de muchii care îl constituie.

- Drum simplu = drum în care nici un vârf nu se repetă.

- Ciclu = drum simplu cu excepția primului și ultimului vârf care coincid.

- Graf aciclic = graf fără cicluri.

Se numește **subgraf** al lui **G** un graf

$$\mathbf{G' = <V', M'>}$$

unde **V'**$\subseteq$ **V**, iar **M'** $\subseteq$ **M** (M' este o mulțime de muchii din M care unesc vârfurile din V').

Se numește **graf parțial** al lui **G** un graf

G” = <V, M”>

unde **M”** $\subseteq$ **M** (graful parțial păstrează aceleași număr de vârfuri **V**).

Alte definiții

- Un graf (orientat sau neorientat) este **conex** dacă între oricare două vârfuri există un drum posibil.

- Un graf orientat este **tare conex** dacă între oricare două varfuri **i** și **j** există un drum de la **i** la **j** și un drum de la **j** la **i**.

- Pentru un graf neconex se pune problema determinării componentelor sale conexe.

- Se numește **componentă conexă** (subgraf conex maximal) al unui graf, un subgraf conex în care nici un vârf din subgraf nu este unit cu unul din afară printr-o muchia a grafului inițial.

Putem atașa informații pentru:

- vârfurile grafului (valori)
- muchii (lungimi sau costuri)

Moduri de reprezentare a unui graf:

a. **Matrice de adiacență** **A**, unde:
- A[i, j] = true ,   dacă i și j sunt adiacente
- A[i, j] = false , dacă i și j nu sunt adiacente

<u>Observație</u>: într-o altă modalitate de reprezentare a matricei de adiacență putem da valori unei muchii oarecare între nodurile i și j, iar dacă respectiva muchie nu există considerăm A[i, j] = ∞ (sau uneori A[i, j] = 0).

Memoria necesară păstrării unei matrice de adiacență ~ $\mathbf{n^2}$

b. **Listă de adiacență**

- în acest caz se atașează fiecărui vârf ***i*** câte o listă de vârfuri adiacente lui ***i*** (pentru un graf orientat este necesar ca muchia să plece din ***i***).

Observație: Pentru un graf cu ***m*** muchii, suma lungimilor listelor de adiacență este:

- **2m** pentru un graf neorientat
- **m** pentru un graf orientat

c. **Listă de muchii**

- aceasta este o reprezentare eficientă când avem de examinat toate muchiile grafului

Arbori

- Definiția 1:

  Arborele este un graf orientat, aciclic și simplu conex.

- Definiția 2:

Un arbore este un ansamblu de structuri de date de natură recursivă şi dinamică.

- Definiția 3:

Prin **arbore** înţelegem o mulţime finită şi nevidă de elemente numite noduri:

ARB = { A1, A2, A3, ..., An } , unde n > 0 ,

care are următoarele proprietăţi:

- Există un nod şi numai unul care se numeşte **rădăcina** arborelui.
- Celelalte noduri formează submulţimi disjuncte ale lui **ARB**, care formează fiecare câte un arbore. Arborii respectivi se numesc **subarbori** ai rădăcinii.

Nodurile unui arbore

- Într-un arbore există noduri cărora nu le mai corespund subarbori. Un astfel de nod se numeşte **terminal** sau **frunză**.
- În legătură cu arborii s-a stabilit un limbaj conform căruia un ***nod rădacină*** se spune că este un ***nod tată***, iar subarborii rădăcinii sunt ***descendenţii*** acestuia.
- ***Rădăcinile descendenţilor*** unui ***nod tată*** sunt ***fiii*** lui.

Nivel

- Rădăcina unui arbore are nivel 1 (sau 0 în alte implementări)

- Dacă un nod se găsește la nivelul **n**, atunci fii lui se găsesc în nivelul **n+1**.

Relația de ordine

- Dacă există o relație de ordine între subarborii oricărui nod al arborelui atunci arborele este **ordonat**.

- Pentru un nod al unui arbore ordonat se notează rădăcina primului subarbore ca fiind fiul cel mai în vârstă, iar rădăcina ultimului subarbore drept fiul cel mai tânăr.

Un **arbore ordonat** este echivalent unui **arbore genealogic** în care rădăcina arborelui este cel mai vechi strămoș cunoscut.

Arbori binari

Un **arbore binar** este o mulţime finită de elemente care sau este **vidă**, sau conţine un element numit **rădăcina**, iar celelalte elemente (dacă există) se împart în două submulţimi disjuncte, care fiecare la rândul ei, este un arbore binar.

Observații:

- Una din submulţimi este numită **subarborele stâng** al rădăcinii, iar cealaltă **subarborele drept**.
- Arborele binar este ordonat, deoarece în fiecare nod, subarborele stâng se consideră că precede subarborele drept.
- De aici rezultă că un nod al unui arbore binar are cel mult doi fii şi că unul este **fiul stâng**, iar celalalt este **fiul drept**.

Observații (continuare)

- Fiul stâng este considerat *mai în vârstă* decât cel drept.

- Un nod al unui arbore binar poate să aibă numai un singur descendent. Acesta poate fi subarborele stâng sau subarborele drept.

Observații (continuare)

- Cele două posibilităţi anterior menționate se consideră distincte.

- Cu alte cuvinte, dacă doi arbori binari diferă numai prin aceea că nodul **V** dintr-un arbore are ca descendent numai fiul stâng, iar acelaşi nod echivalent, din celălalt arbore, are ca descendent numai fiul drept, cei doi arbori se consideră distincţi.

Particularități

- Un arbore binar nu se defineşte ca un caz particular de arbore ordonat. Astfel, un arbore nu este niciodată vid, spre deosebire de un arbore binar care poate fi şi vid.

- **Orice arbore ordonat poate fi întotdeauna reprezentat printr-un arbore binar.**

Modalitatea de transformare a unui arbore ordonat într-un arbore binar

1. Se leagă între ei frații descendenți ai aceluiași nod tată și se suprimă legăturile lor cu nodul tată, cu excepția primului fiu.
2. Fostul prim fiu devine fiul stâng al nodului tată, iar ceilalți foști frați formează, în mod consecutiv, subarbori drepți. Fiecare dintre foștii frați devine descendent drept (fiu drept) al fostului frate mai mare.

Utilizarea structurilor pentru realizarea unui arbore binar

Nodul unui arbore binar poate fi reprezentat ca o dată structurală de tipul NOD care se defineşte în felul următor:

```
typedef struct nod
    {
        declaratii ;
        struct nod *st;
        struct nod *dr;
    }  NOD;
```

unde:

- st - este pointerul spre fiul stâng al nodului curent;
- dr - este pointerul spre fiul drept al aceluiaşi nod.

Asupra arborilor binari se pot defini mai multe operaţii:

1. inserarea unui nod frunză într-un arbore binar;
2. accesul la un nod al unui arbore;
3. parcurgerea unui arbore;
4. ştergerea unui arbore.

- Operaţiile de **inserare** şi **acces** la un nod, au la bază un **criteriu** care să definească locul în arbore al nodului în cauză.

- Acest **criteriu** este dependent de problema concretă, la care se aplică arborii binari, pentru a fi rezolvată.

Funcția criteriu

Definim o funcţie pe care o vom denumi **criteriu**. Aceasta are doi parametri care sunt pointeri spre tipul NOD.

Fie p1 primul parametru al funcţiei **criteriu** şi p2 cel de-al doilea parametru al ei. Atunci, funcţia **criteriu** returnează:

- **-1** - dacă p2 pointează spre o dată de tip NOD care poate fi un nod al subarborelui stâng al nodului spre care pointează p1;
- **1** - dacă p2 pointează spre o dată de tip NOD care poate fi un nod al subarborelui drept al nodului spre care pointează p1;
- **0** - dacă p2 pointează spre o dată de tip NOD care nu poate fi nod al subarborilor nodului spre care pointează p1.

Suplimentar

La construirea unui arbore se stabileşte un criteriu pentru determinarea poziţiei în care să se insereze nodul curent – adică nodul corespunzător valorii (sau uneori grupului de valori) curent achiziționate.

Exemplu:

Considerăm următorul șir de numere:

20, 30, 5, 20, 4, 30, 7, 40, 25, 28, ...

Să se creeze un arbore în nodurile căruia vor fi trecute numerele respective împreună cu frecvența lor de apariție.

Practic nodurile acestui arbore vor avea două câmpuri utile:

- primul care va conține numărul;

- celălalt care va conține frecvența de apariție a numărului.

Abordarea problemei:

a. **p1** - este pointer spre un nod al arborelui în care se face inserarea (iniţial p1 pointează spre rădacina arborelui).

b. **p2** - este un pointer spre nodul curent (nodul de inserat).

c. dacă **p2->val < p1->val**, atunci se va încerca inserarea nodului curent în subarborele stâng al nodului spre care pointează p1.

d. dacă **p2->val > p1->val**, atunci se va încerca inserarea nodului curent în subarborele drept al nodului spre care pointează p1.

e. dacă **p2->val = p1->val**, atunci nodul curent nu se mai inserează în arbore deoarece există deja un nod corespunzător valorii citite curent.

Nodul curent nu se mai inserează în arbore în cazul descris la punctul **e** (situaţie care în general corespunde cazului când funcţia **criteriu** returnează valoarea zero). În acest caz, nodurile spre care pointează p1 şi p2 le vom considera **echivalente**.

În cazul exemplului precedent, funcția criteriu este:

```
int criteriu(NOD *p1, NOD *p2)
{
        if(p2->nr < p1->nr) return(-1);
         if(p2->nr > p1->nr) return(1);
        return(0);
}
```

Funcția pentru tratarea echivalenței

- De obicei, la întâlnirea unei perechi de noduri echivalente, nodul din arbore (spre care pointează p1) este supus unei prelucrări, iar nodul curent (spre care pointează p2) este eliminat.
- Pentru a realiza o astfel de prelucrare este necesar să se apeleze o funcţie care are ca parametri pointerii p1 şi p2 şi care returnează un pointer spre tipul NOD (de obicei se returnează valoarea lui p1).
- Vom numi această funcţie: **echivalenta**. Ea este dependentă de problema concretă, ca şi funcţia **criteriu**.

Exemplu:

```
NOD *echivalenta(NOD *q, NOD *p)
{
        q -> frecventa ++;
        elibnod(p);
        return(q);
}
```

Observație:

Funcția de mai sus realizează următoarele:

- eliberează zona de memorie ocupată de nodul spre care pointează **p**;
- incrementează valoarea componentei: **q -> frecventa**;
- returnează valoarea lui **q**.

Alte funcții

- În afară de funcţiile enumerate anterior, vom folosi nişte funcţii specifice pentru operaţii asupra arborilor.
- Exemple tipice de funcții: **elibnod** și **incnod**
- Unele funcții utilizează o variabilă globală care este un pointer spre rădăcina arborelui.

Exemplu de funcție folosită la laborator:

```
void elibnod(NOD *p)
/* elibereaza zonele din memoria heap ocupate de
nodul spre care pointează p */
{
 free(p -> cuvant);
 free(p);
}
```

Intrarea în arbore

- Numim **prad** o variabilă globală către rădăcina arborelui binar. Ea se defineşte astfel:

  NOD *prad;

- În cazul în care într-un program se prelucrează simultan mai mulţi arbori, se vor utiliza mai multe variabile globale de tip **prad**; interfaţa dintre funcţii realizându-se cu ajutorul unui parametru care este pointer spre tipul NOD şi căruia i se atribuie, la apel, adresa nodului rădăcină al arborelui prelucrat prin funcţia apelată.

În continuare vom folosi funcţii care utilizează variabila globală **prad**.

1. Inserarea unui nod frunză într-un arbore binar

Funcţia **insnod** (declarată NOD *insnod()) inserează un nod nou **p** în arbore, conform următorilor paşi:

1. Se alocă zona de memorie pentru nodul care urmează să se insereze în arbore. Fie **p** pointerul care are ca valoare adresa de început a zonei respective.

2. Se apelează funcţia **incnod**, cu parametrul **p**, pentru a încărca datele curente în zona spre care pointează p. Dacă **incnod** returnează valoarea 1, se trece la pasul 3. Altfel se revine din funcţie cu valoarea zero.

3. Se fac atribuirile:

p->st = p->dr = 0

deoarece nodul de inserat este nod frunză.

4. **q = prad**

5. Se determină poziţia, în arbore, în care sa se faca inserarea. În acest scop se caută nodul care poate fi nod tată pentru nodul curent:

i = criteriu(q,p)

6. Dacă **i<0**, se trece la pasul 7; altfel se sare la pasul 8.

7. Se încearcă inserarea nodului spre care pointează **p** (nodul curent) în subarborele stâng al arborelui spre care pointează **q**.
- Dacă **q -> st** are valoarea zero, atunci nodul spre care pointează q nu are subarbore stâng şi nodul curent devine fiu stâng al celui spre care pointează q (**q->st = p**). Se revine din funcţie returnându-se valoarea lui p.
- Altfel se face atribuirea **q = q->st** (se trece la fiul stâng al nodului spre care pointează **q**) şi se sare la pasul 5.

8. Dacă **i>0**, se trece la pasul 9; altfel se sare la pasul 10.

9. Se încearcă inserarea nodului spre care pointează **p** (nodul curent) în subarborele drept al arborelui spre care pointează **q**.
- Dacă **q -> dr** are valoarea zero, atunci nodul spre care pointează q nu are subarbore drept şi nodul curent devine fiu drept al celui spre care pointează **q** (**q->dr=p**). Se revine din funcţie returnându-se valoarea lui p.
- Altfel se face atribuirea **q = q->dr** (se trece la fiul drept al nodului spre care pointează **q**) şi se sare la pasul 5.

10. Nodul curent nu poate fi inserat în arbore. Se apelează funcţia numită **echivalenta** şi se revine din funcţie cu valoarea returnată de funcţia **echivalenta**.

Exemplu:

Reamintire $\rightarrow$ șirul de numere:

20, 30, 5, 20, 4, 30, 7, 40, 25, 28, ...

2. Accesul la un nod al unui arbore binar

- Accesul la un nod presupune existența unui criteriu care să permită localizarea în arbore a nodului respectiv.

- Se va utiliza funcția **criteriu**.

- Funcția care realizează căutarea va identifica în arbore un nod echivalent cu cel spre care pointează un pointer **p** (folosit ca argument al funcției de căutare).

Funcția de căutare (numită **cauta**) va returna:

- pointerul spre nodul determinat;

- 0 dacă nu există un nod echivalent cu **p**.

```
NOD *cauta(NOD *p)
{
        extern NOD *prad;
        NOD *q;
        int i;
        if (prad = = 0) return 0;  /*arborele este vid*/
        for (q = prad; q;  )
          {
                if ( (i = criteriu(q, p)) = = 0)    return q;
                else   if (i < 0)        q = q -> st;
                       else              q = q -> dr;
          }
        return 0;
}
```

3. Parcurgerea unui arbore binar

- Prelucrarea informaţiei păstrată în nodurile unui arbore binar se realizează parcurgând nodurile arborelui respectiv.

- Parcurgerea nodurilor unui arbore binar se poate face în mai multe moduri, dintre care se remarcă:
  – **parcurgerea în preordine;**
  – **parcurgerea în inordine;**
  – **parcurgerea în postordine.**

Parcurgerea în preordine

Parcurgerea în preordine înseamnă accesul la rădăcină şi apoi parcurgerea celor doi subarbori ai ei, întâi subarborele stâng, apoi cel drept. Subarborii, fiind ei înşişi arbori binari, se parcurg în acelaşi mod.

Parcurgerea în inordine

Parcurgerea în inordine înseamnă parcurgerea mai întâi a subarborelui stâng, apoi accesul la rădăcină şi în continuare parcurgerea subarborelui drept. Cei doi subarbori se parcurg în acelaşi mod.

Parcurgerea în postordine

Parcurgerea în postordine înseamnă parcurgerea mai întâi a subarborelui stâng, apoi a subarborelui drept şi în final accesul la rădăcina arborelui. Cei doi subarbori se parcurg în acelaşi mod.

- Accesul la un nod permite prelucrarea informaţiei conţinute în nodul respectiv.

- În acest scop se poate apela o funcţie care este dependentă de problema concretă care se rezolvă cu ajutorul parcurgerii arborelui (de exemplu funcţia **prelucrare**).

```
void prelucrare( NOD *p)
```

Parcurgerea unui arbore binar in preordine

```
void preord(NOD *p)
{
 if(p != 0)
       {
        prelucrare(p);
        preord(p -> st);
        preord(p -> dr);
       }
}
```

Parcurgerea unui arbore binar in inordine

```
void inord(NOD *p)
{
 if(p != 0)
       {
        inord(p -> st);
        prelucrare(p);
        inord(p -> dr);
       }
}
```

Parcurgerea unui arbore binar in postordine

```
void postord (NOD *p)
{
 if(p != 0)
       {
        postord (p -> st);
        postord (p -> dr);
        prelucrare(p);
       }
}
```

Exemplu:

Reluăm problema cu șirul de numere:

20, 30, 5, 20, 4, 30, 7, 40, 25, 28, ...

Observații:

- Programul de la laborator (cu noduri care conțin cuvinte împreună cu frecvența lor de apariție într-un text) se rezolvă mai ușor prin intermediul arborilor (deoarece operația de căutare într-o listă este ineficientă).

- Procesul de căutare într-un arbore necesită mai puțini pași decât procesul de căutare într-o listă.

4. Ștergerea unui arbore binar

- pentru ștergerea unui arbore binar este necesară parcurgerea lui și ștergerea fiecărui nod al arborelui respectiv.

- se apelează funcția elibnod.

- arborele se parcurge în postordine

Ștergerea unui arbore binar in postordine

```
void sterge_arb(NOD *p)
{
 if(p != 0)
       {
        sterge_arb (p -> st);
        sterge_arb (p -> dr);
        elibnod(p);
       }
}
```

Observații:

- Funcția **sterge_arb** nu atribuie valoarea zero variabilei globale prad.

- Această atribuire se va face obligatoriu imediat după ce se apelează funcția **sterge_arb**

- **Arbore binar degenerat** – dacă și numai dacă toți subarborii lui sunt de același fel (adică sunt numai stângi, sau numai drepți).

5. Ștergerea unui nod precizat printr-o cheie

- Cheia după care se face căutarea este unică și se găsește în fiecare nod:

```
typedef struct nod
        {
                declaratii ;
                tip cheie;
                struct nod *st;
                struct nod *dr;
        }  NOD;
```

unde tipul cheii poate fi char, int, float sau double.

- **Se pot șterge doar noduri care sunt frunză**!

- Ștergerea unui nod care nu este frunză implică operații suplimentare complicate pentru refacerea arborelui.

```
void cauta_sterge(NOD *p, int c)    /*cheia este de tip int*/
{
    if (p != 0)
      {
        if (( p -> st = = p -> dr ) && ( p -> st = = 0 ) && ( p -> cheie = = c))
                {
                  elibnod (p);
                  return;
                }
         cauta_sterge(p -> st, c);
         cauta_sterge(p -> dr, c);
      }
}
```

Observație: această operație de căutare și ștergere se face în **preordine**.

Observație:

- La o analiză atenta, funcției anterioare îi lipsește ceva: după ce o un nod frunză identificat a fost șters, tatăl său ar trebui să aibă pointerul recursiv zero catre direcția proaspăt ștearsă!

- Cum rezolvați această problemă?

Precizare

- În anumite aplicații sunt necesari arbori mai aplatizați, de înălțime mică, dar cu număr mare de noduri.
- În aceste cazuri se pot înlocui arborii binari cu unii care să permită fiecărui nod un număr mai mare de descendenți direcți.
- Procedura este relativ simplă, iar în loc de cei doi pointeri recursivi (către subarborele stâng și către cel drept), se utilizează un vector de pointeri recursivi (cu o anumită lungime maximă) care permite tipului de structură, utilizată pentru definirea nodurilor, un număr arbitrar de mare de descendenți direcți.