

# Struktura indeksa: B-stablo

<http://cis.stvincent.edu/html/tutorials/swd/btree/btree.html>

# Uvod

- ISAM (Index-Sequential Access Method, IBM sredina 60-tih godina 20. veka)
- Nedostaci: sekvencijalno pretraživanje indeksa
- Poboljšanje: indeks nad indeksom
- Uopštavanje: drvoidna indeksna struktura
  - svaki nivo osim nivoa lista je redak indeksa
  - ilustracija

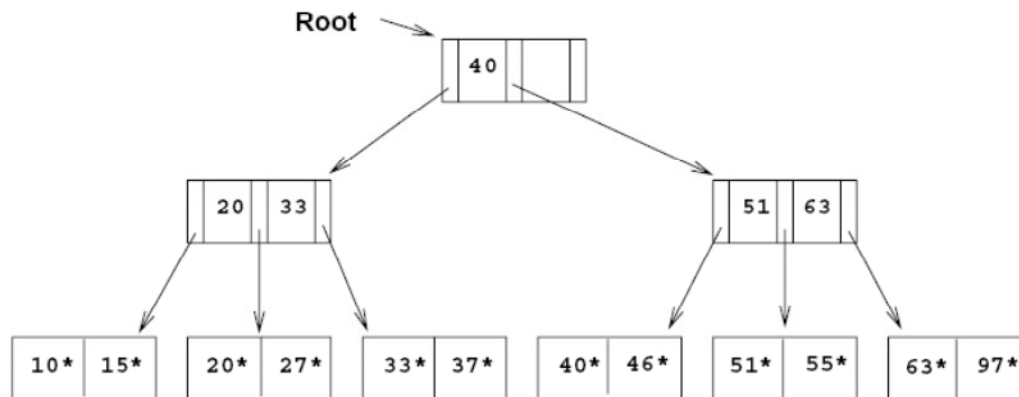


Figure 9.5 Sample ISAM Tree

# ISAM

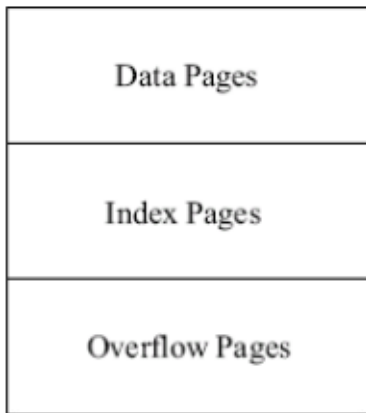


Figure 9.4 Page Allocation in ISAM

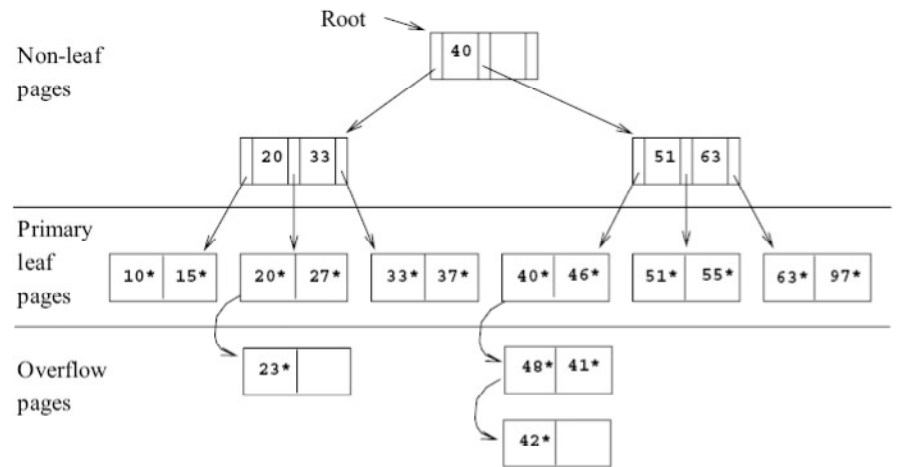


Figure 9.6 ISAM Tree after Inserts

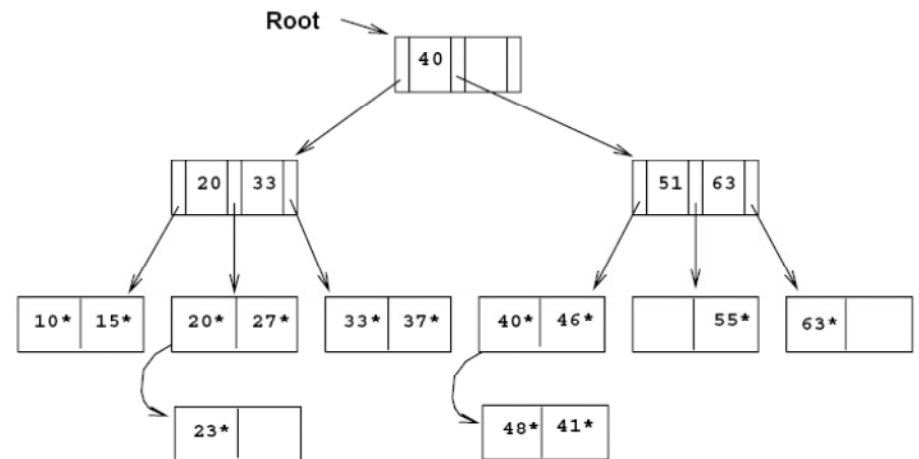


Figure 9.7 ISAM Tree after Deletes

# Definicije

## Graf

- *Usmereni graf*  $\Gamma$  je uređeni par  $(A, r)$ , gde je  $A$  – konačni skup a  $r$  – binarna relacija nad skupom  $A$ .
- Elemente skupa  $A$  zovemo *čvorovi*, a elemente skupa  $r$  – *potezi*. Ako je  $(a_i, a_j) \in r$  poteg, kaže se da je čvor  $a_i$  – *neposredni prethodnik* čvora  $a_j$ , a da je čvor  $a_j$  – *neposredni sledbenik* čvora  $a_i$ ; takođe, čvor  $a_i$  je *izlazni čvor* potega  $(a_i, a_j)$ , a čvor  $a_j$  je *ulazni čvor* potega  $(a_i, a_j)$ .
- *Ulazni red* čvora  $a_i$  je broj potega kojima je taj čvor – ulazni čvor, a *izlazni red* čvora  $a_i$  je broj potega kojima je taj čvor – izlazni čvor.
- *Putanja dužine  $n$*  od čvora  $a_i$  do čvora  $a_k$  u grafu  $\Gamma$  je niz čvorova  $(a_i, a_{i+1}, \dots, a_{i+n})$ , takvih da je  $a_{i+n} = a_k$  i  $(a_j, a_{j+1}) \in r$  za svako  $j = i, i + 1, \dots, i+n-1$ . Putanja od čvora  $a_i$  do čvora  $a_k$ , dužine veće od 1, za koju važi da je  $a_i = a_k$ , zove se *ciklus*.

# Stablo

## Stablo

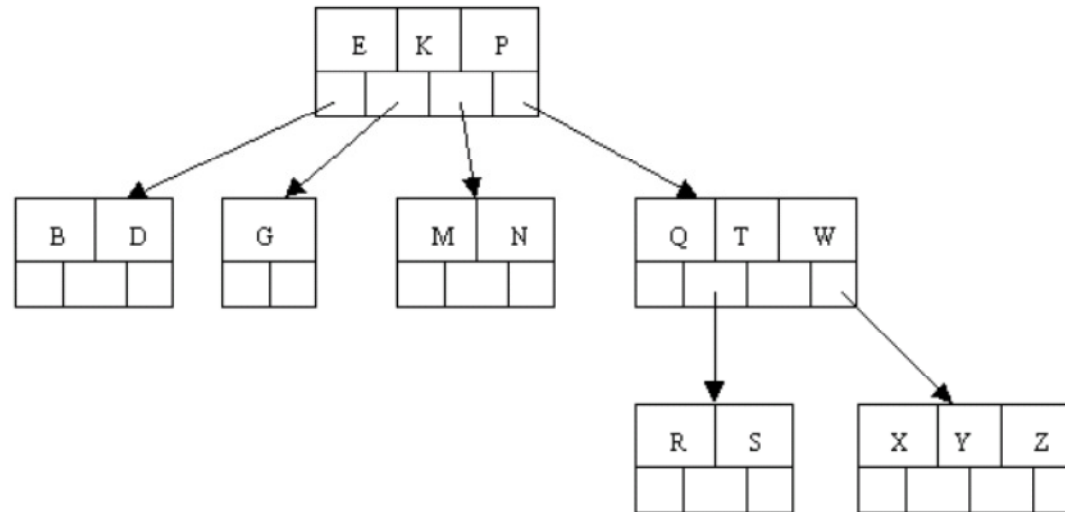
Graf  $\Gamma$  je *stablo* ako zadovoljava sledeće uslove:

- ulazni red svakog čvora (osim jednog) je 1;
  - ulazni red jednog čvora je 0, i taj čvor se zove *koren* stabla;
  - izlazni red svakog čvora je  $\geq 0$  i konačan;
  - graf  $\Gamma$  ne sadrži cikluse.
- 
- čvorovi bez neposrednih sledbenika zovu se *listovi* stabla (izlazni red svakog lista je 0);
  - *visina* stabla je maksimalna dužina putanje od korena do nekog lista stabla.
  - čvorovi – neposredni sledbenici istog čvora zovu se *braća*

# Stablo reda $m$

- Stablo reda  $m$  je stablo u kome je izlazni red svakog čvora manji ili jednak  $m$  (svaki čvor ima najviše  $m$  sledbenika).
- Svaki čvor sadrži određen broj *ključeva*; čvor sa  $k$  sledbenika sadrži tačno  $k - 1$  ključeva.
- Ako su ključevi i podstabla uređeni kao kod stabla pretrage, stablo reda  $m$  naziva se **stablom pretrage reda  $m$** .
- Primer: stablo pretrage reda 4 (prvi red čvora prikazuje ključeve, drugi red pokazivače na čvorove sledbenike).
- U realnoj situaciji, umesto ključa biće ceo slog, pa će u prvom redu biti niz slogova koji uključuje ključ i pridružene podatke (ili ključ i redni broj sloga, ako su slogovi smešteni u posebnoj datoteci).

# Stablo pretrage reda m: primer



# Stablo pretrage reda m: osobine

- Čvor stabla pretrage reda m može da se predstavi strukturom sa brojem ključeva (*int Broj*) i dva niza: nizom *Ključevi* od m-1 ključeva nekog tipa i nizom *Grane* od m pokazivača na strukture istog tipa
- Struct NodeType  
{  
    int Broj; // broj ključeva smeštenih u tekućem čvoru  
    tipKljuča Ključevi[m-1]; // niz za najviše m-1 ključeva  
    struct NodeType \*Grane [m]; // niz pokazivača  
} N;
- U tom slučaju, stablo pretrage reda 4 (slično za m) treba da zadovolji sledeće uslove u vezi sa uređenjem ključeva:



# Stablo pretrage reda m: osobine

- Ključevi u svakom čvoru su u rastućem poretku.
- U svakom čvoru N važi:
  - Podstablo koje počinje u strukturi N.Grane[0] sadrži ključeve koji su manji od N.Ključevi[0].
  - Podstablo koje počinje u strukturi N.Grane[1] sadrži ključeve koji su veći od N.Ključevi[0] i istovremeno manji od N.Ključevi[1].
  - Podstablo koje počinje u strukturi N.Grane[2] sadrži ključeve koji su veći od N.Ključevi[1] i istovremeno manji od N.Ključevi[2].
  - Podstablo koje počinje u strukturi N.Grane[3] sadrži ključeve koji su veći od N.Ključevi[2].
- Ne govorimo o duplikatima

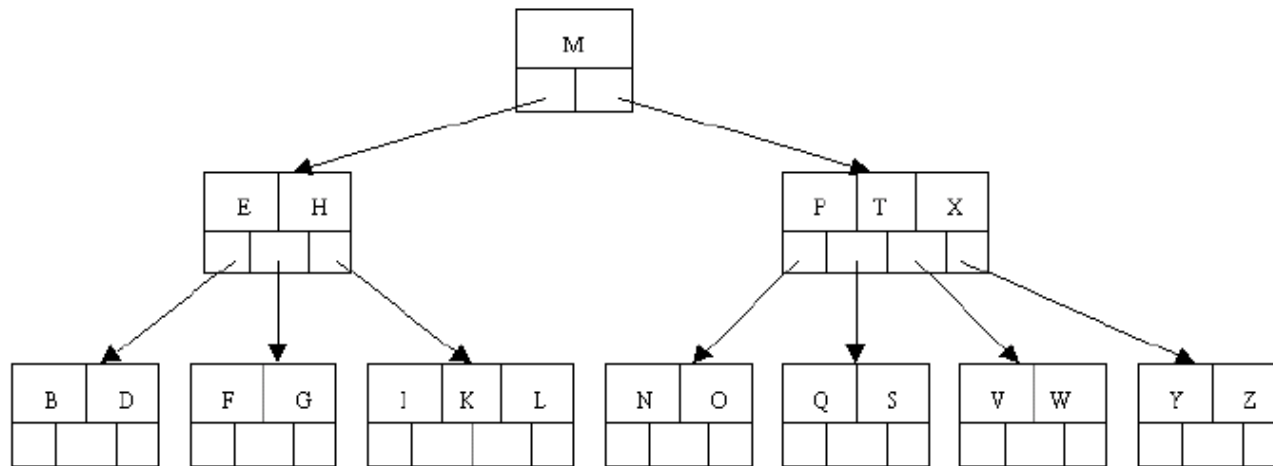
# B-stablo

- Posebno poželjna osobina stabla pretrage reda  $m$ , kada je u pitanju efikasnost njegovog pretraživanja, jeste *balansiranost* koja označava da je dužina puta od korena do svakog lista jednaka.
- Visina stabla:  $\log_m Br$ ,  $Br$  – ukupan broj ključeva u stablu; obično 3-4 (npr. 10000000 ključeva,  $m=50$ , visina 4)
- Jedna specifična struktura balansiranog stabla, poznata kao **B-stablo**, poslužila je kao osnov za izgradnju niza struktura podataka koje se koriste u implementaciji indeksa.
- Struktura **specijalizovana za korišćenje na disku**
- **Čvorovi** – blokovi fiksne veličine na disku – “**stranice**”
- Puno ključeva na jednoj stranici, puno sledbenika, veliki red  $m$  – velika efikasnost pristupa podacima na disku

# B-stablo

- B-stablo reda  $m$  je stablo pretrage reda  $m$  takvo da:
  - Svi listovi su na najnižem nivou (stablo je balansirano)
  - Svi unutrašnji čvorovi (osim možda korena) imaju najmanje  $\lceil m / 2 \rceil$  (nepraznih) sledbenika.
  - Koreni čvor, ako nije list, ima najmanje 2 sledbenika, a ako je istovremeno i list onda nema sledbenika (i stablo se sastoji samo od tog jednog čvora)
  - Svaki list sadrži najmanje  $\lceil m / 2 \rceil - 1$  ključeva
- B-stablo je osnovna drvoidna indeksna struktura
- Sadrži podatke (ključeve tj. slogove) **na svim nivoima**
- Popunjenost stranice min 50%, u proseku oko 70%
- Modifikacije: B\* stablo, B+ stablo, sa većom popunjenošću / razdvojenim indeksom od podataka / sekvencijalnošću nad podacima

# B-stablo: primer - red 5



# Operacije nad B-stablom: pretraživanje

- (1) Neka je stranica  $N$  - korena stranica
- (2) Pretražiti stranicu  $N$  na ključ  $Klj$
- (3) Ako je ključ  $Klj$  nađen, pretraživanje uspešno
- (4) Inače,
  - Ako je stranica  $N$  list, pretraživanje je neuspešno
  - Inače, naći među ključevima  $N.Kljucevi[0]$ ,  $N.Kljucevi[1]$ ,  $N.Kljucevi[2]$ , ...,  $N.Kljucevi[Br-1]$  najmanji ključ koji je veći od  $Klj$ ; neka je to ključ  $N.Kljucevi[j]$ ;  $N$  postaje stranica na koju pokazuje pokazivač  $N.Grane[j]$ ; Ako takav ključ ne postoji,  $N$  postaje stranica na koju pokazuje pokazivač  $N.Grane[Br]$ ; preći na korak 2.
- Primer: pretražiti prethodno stablo na ključeve  $S$ ,  $J$ ,  $A$

- <http://www.csanimated.com/animation.php?t=B-tree>
- <http://wiki.algoviz.org/AlgovizWiki/B-Trees>
- [http://www.cs.hut.fi/Research/TRAKLA2/exercises/BTree\\_insert.html](http://www.cs.hut.fi/Research/TRAKLA2/exercises/BTree_insert.html)