

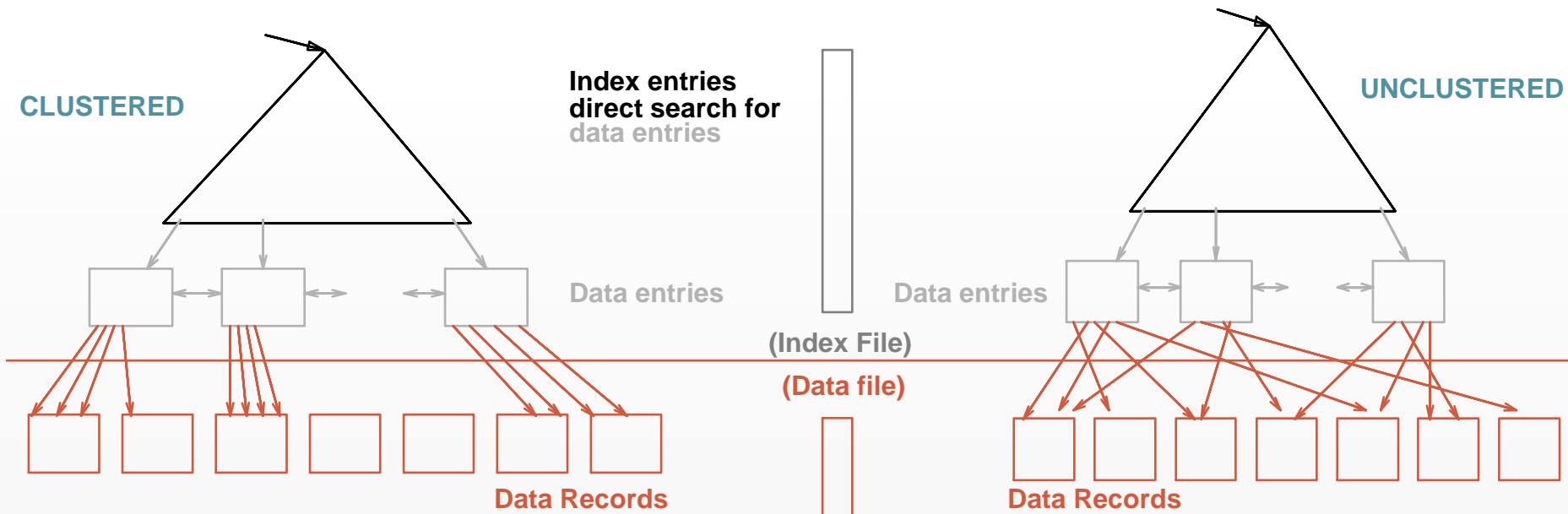
Uređeni indeksi

Uređeni indeksi

- Struktura podatka u kojoj su parovi key, data entry uređeni prema vrednosti ključa.
Key - može biti jedan ili više atributa tabele nad kojima je postavljen indeks.
Data entry – ukazuje na torku u fajlu (tabeli).
 - Struktura se beleži tako da omogući brzu pretragu.
 - DBMS ima obavezu da pri izmeni sadržaja tabele sinhronizovano vrši i izmene u pridruženim indeksima.
 - Tokom izvršavanja upita DBMS procenjuje koje indekse je moguće iskoristiti za ubrzavanje izvršavanja, tj. smanjenje broja čitanja/pisanja po disku.
 - Troškovi:
 - Memorijski prostor
 - Održavanje
-

Primarni i sekundarni indeksi

- **Primarni indeks.** Indeksirano polje odgovara kriterijumu po kojem je fajl sa podacima uređen.
- **Sekundarni indeks.** Indeks u kojem indeksirano polje nameće uređenje koje ne odgovara sekvencijalnom uređenju fajla.



Primarni indeksi

- Indeks-sekvencijalna fajl organizacija.

Sekvencijalni fajl sa primarnim indeksom.

- Gusti indeks.

U indeksu postoji slog za svaku vrednost indeksiranog polja.

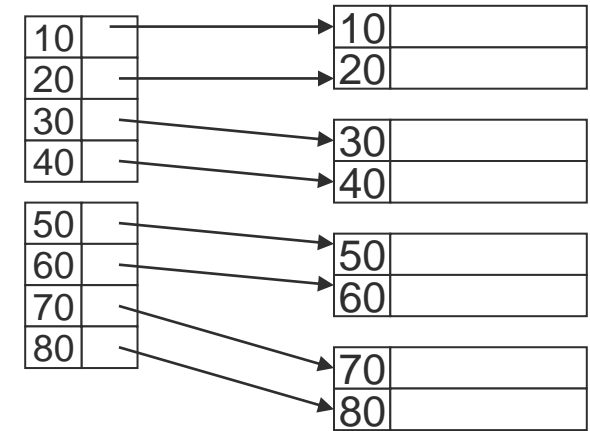
- Redak indeks.

U indeksu postoje samo neke vrednosti indeksiranog polja.

Obično sadrže jedan ključ-pokazivač par po bloku.

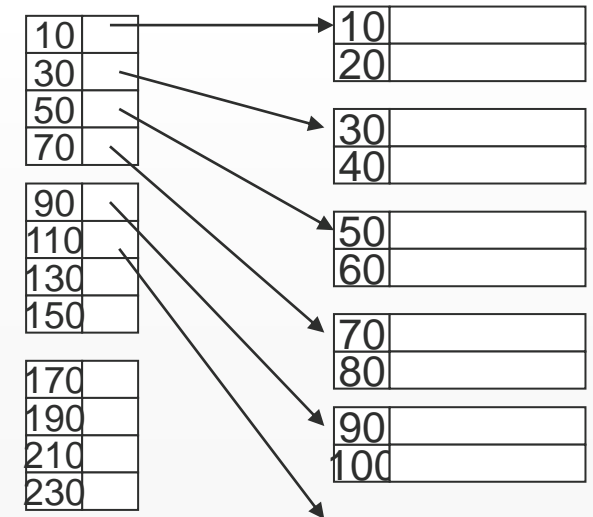
Primenljiv samo u slučaju sekvencijalnog uređenja prema indeksiranom polju.

Dense Index



Sparse Index

Sequential File



Gusti vs. Retki

- **Gusti indeksi.**

Može se utvrditi da li neka vrednost indeksiranog polja postoji bez učitavanja fajla sa podacima.

Zauzima više prostora.

Sporije se pretražuje sam indeks.

- **Retki indeksi.**

Veći deo indeksa može biti učitano u memoriju.

Lakši za implementaciju dodavanja novih slogova.

Duplikati / gust indeks

10	
10	

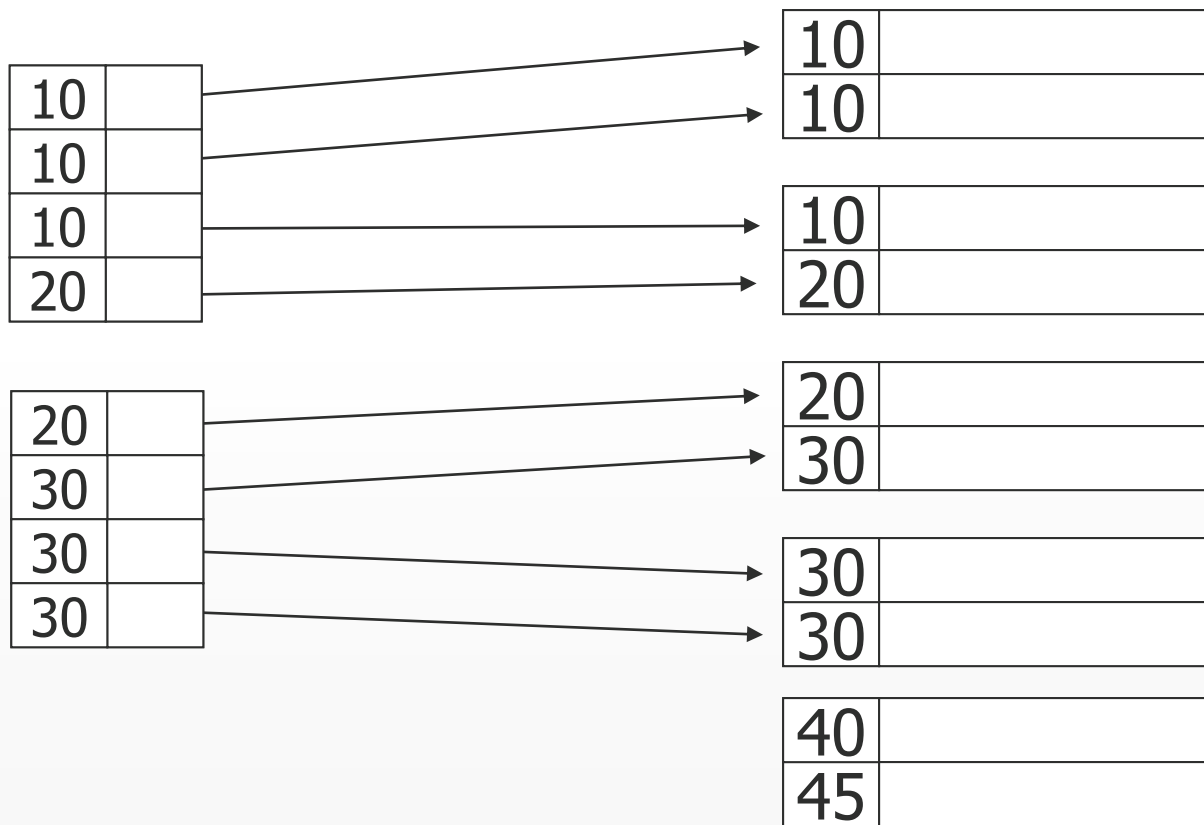
10	
20	

20	
30	

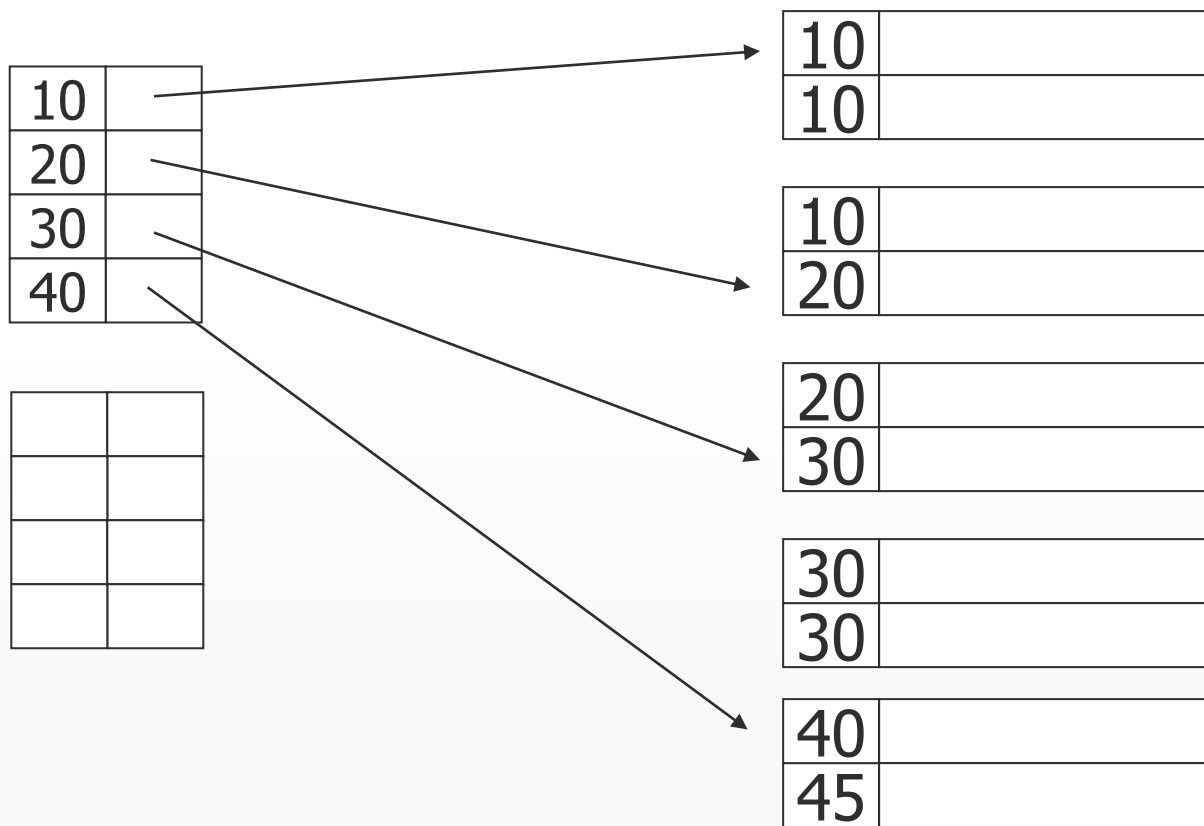
30	
30	

40	
45	

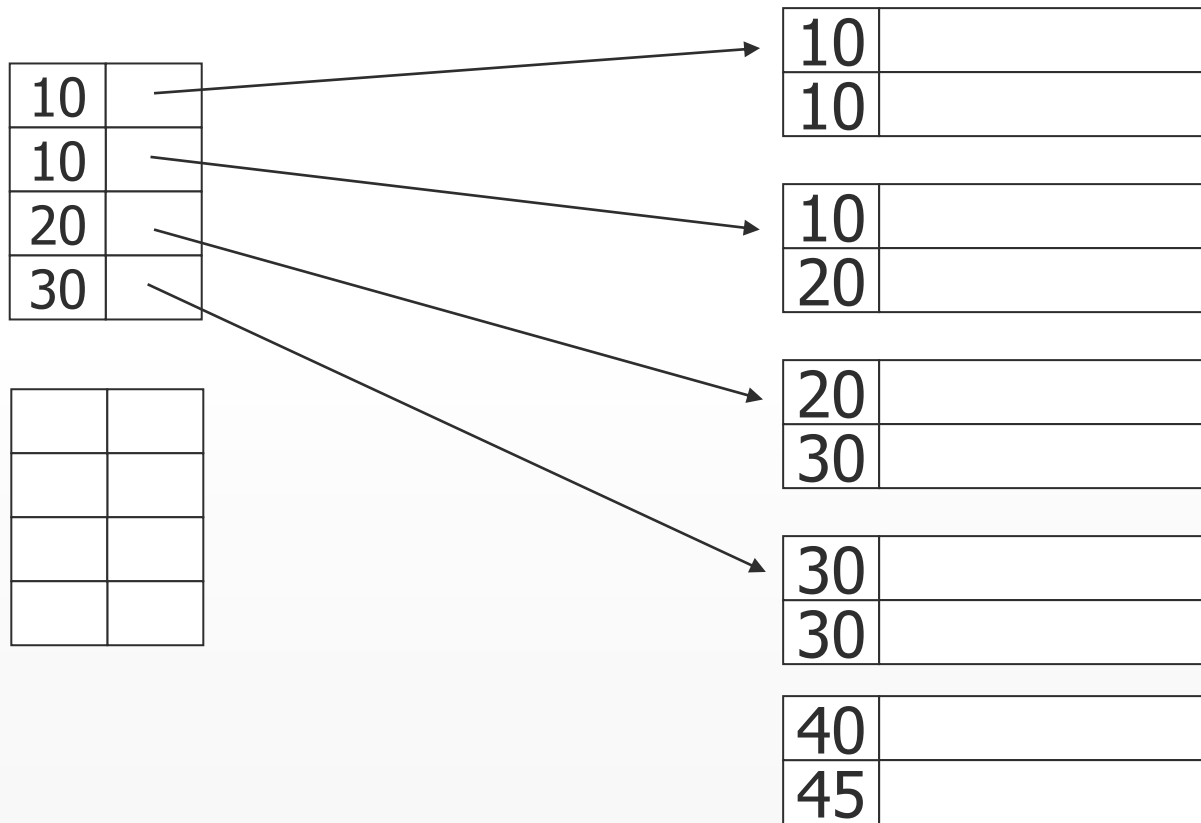
Duplikati / gust indeks – prvi način



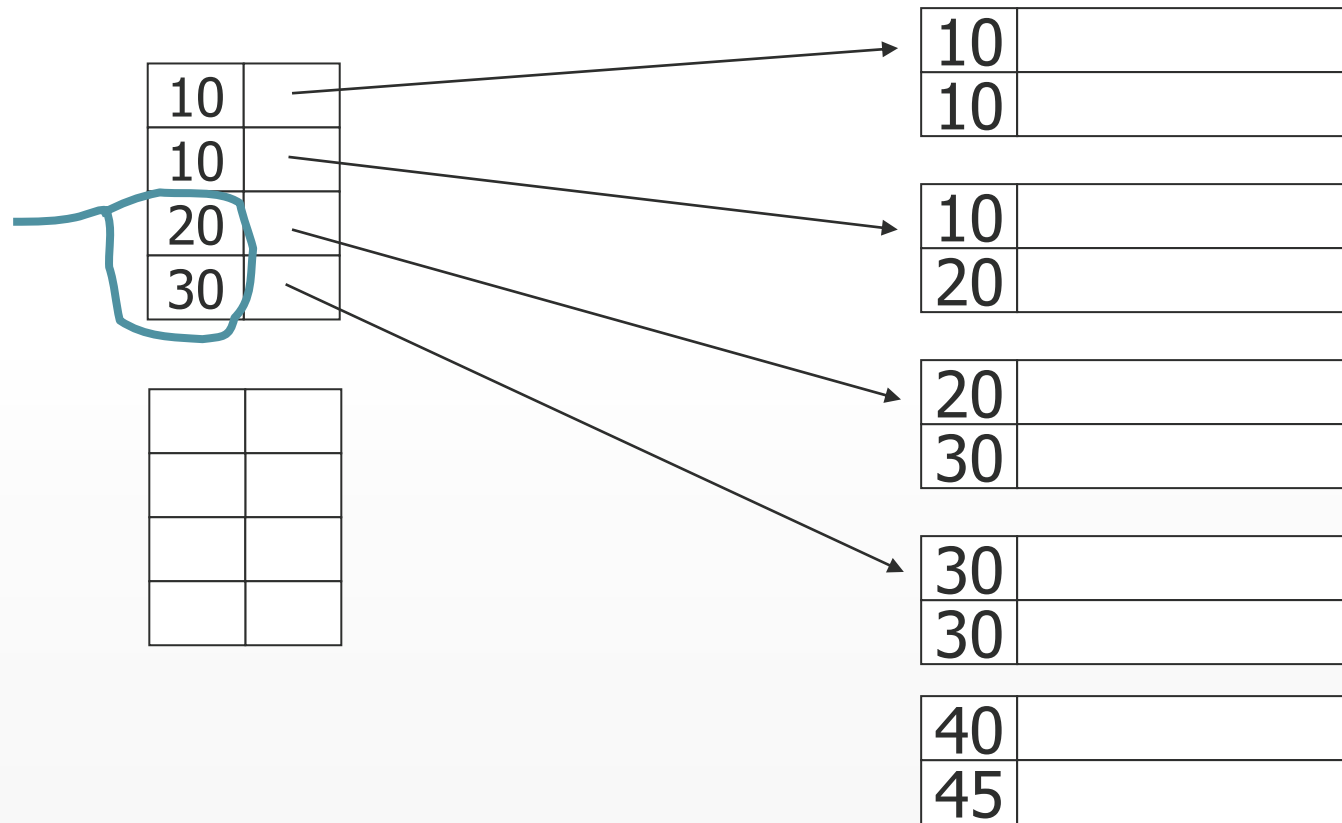
Duplikati / gust indeks – drugi način



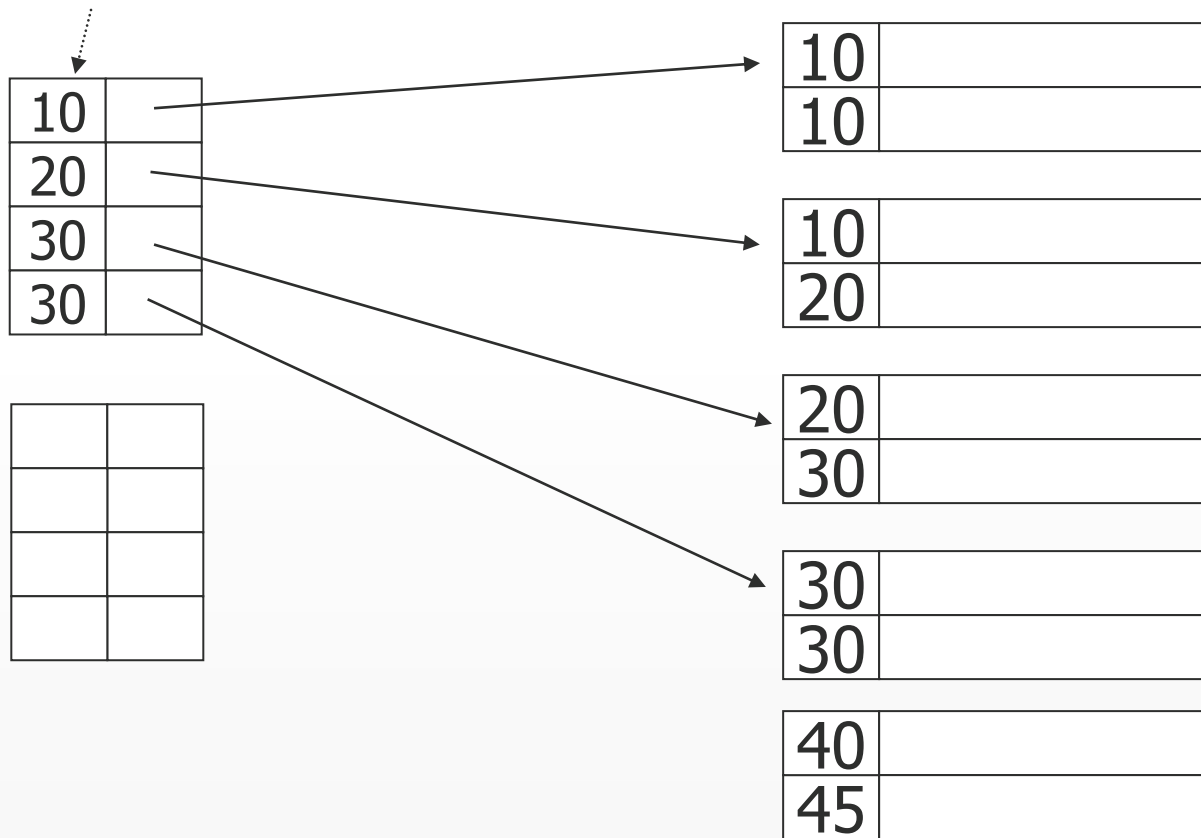
Duplikati / redak



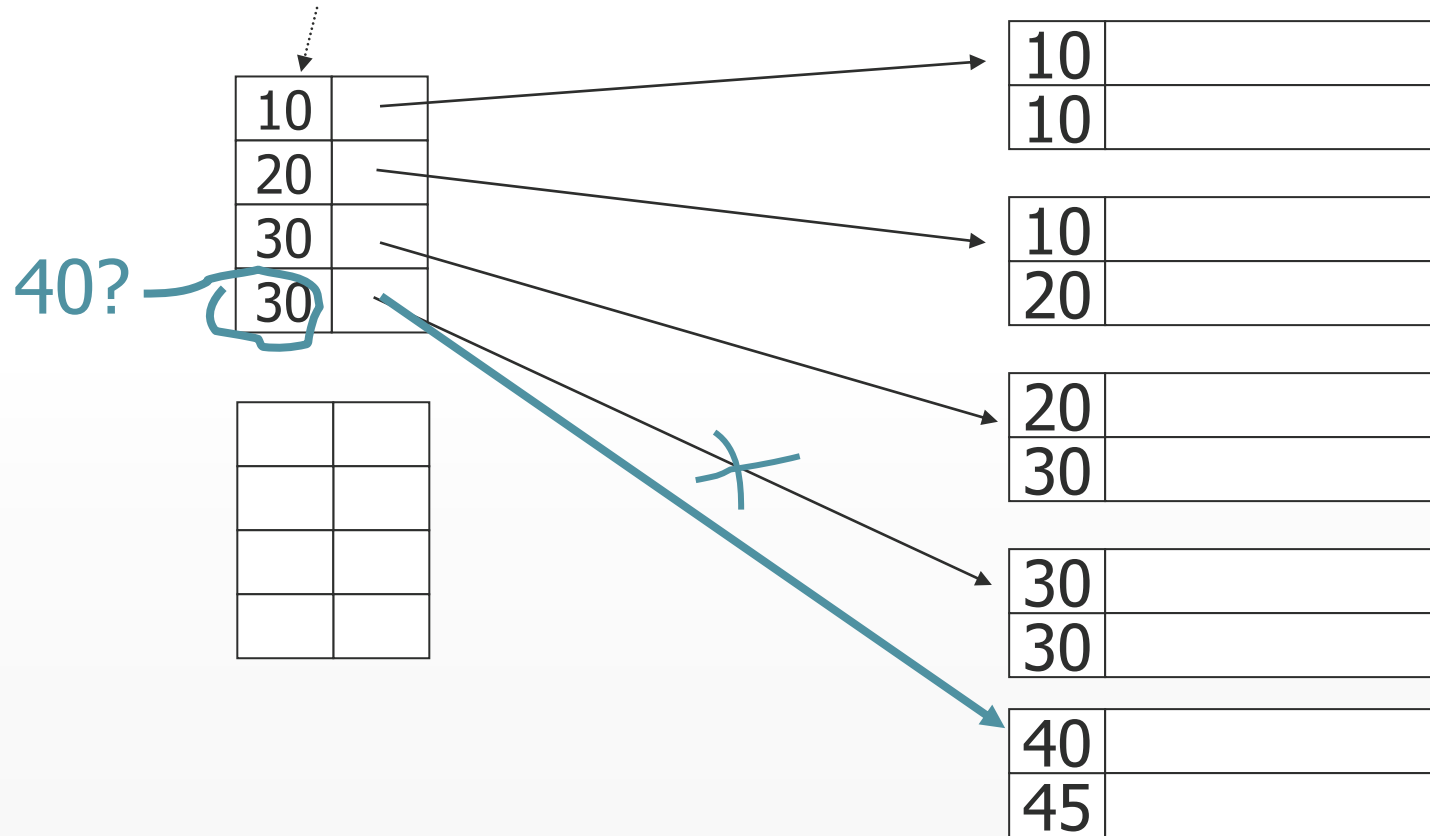
Duplikati / redak



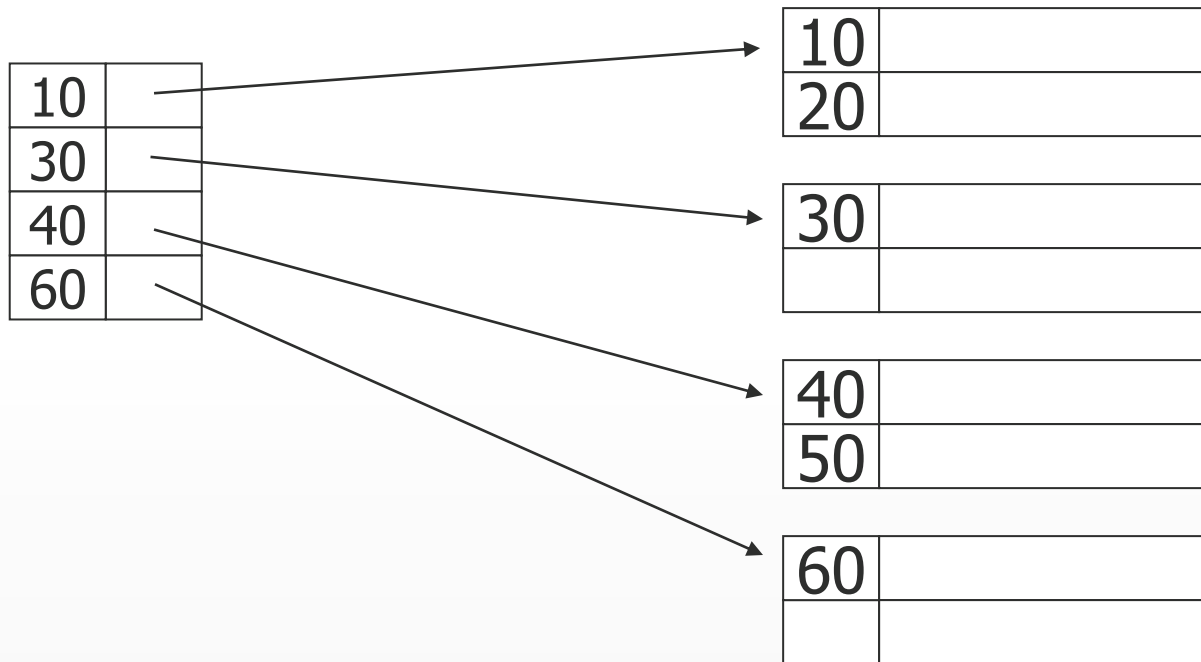
Duplikati / redak



Duplikati / redak

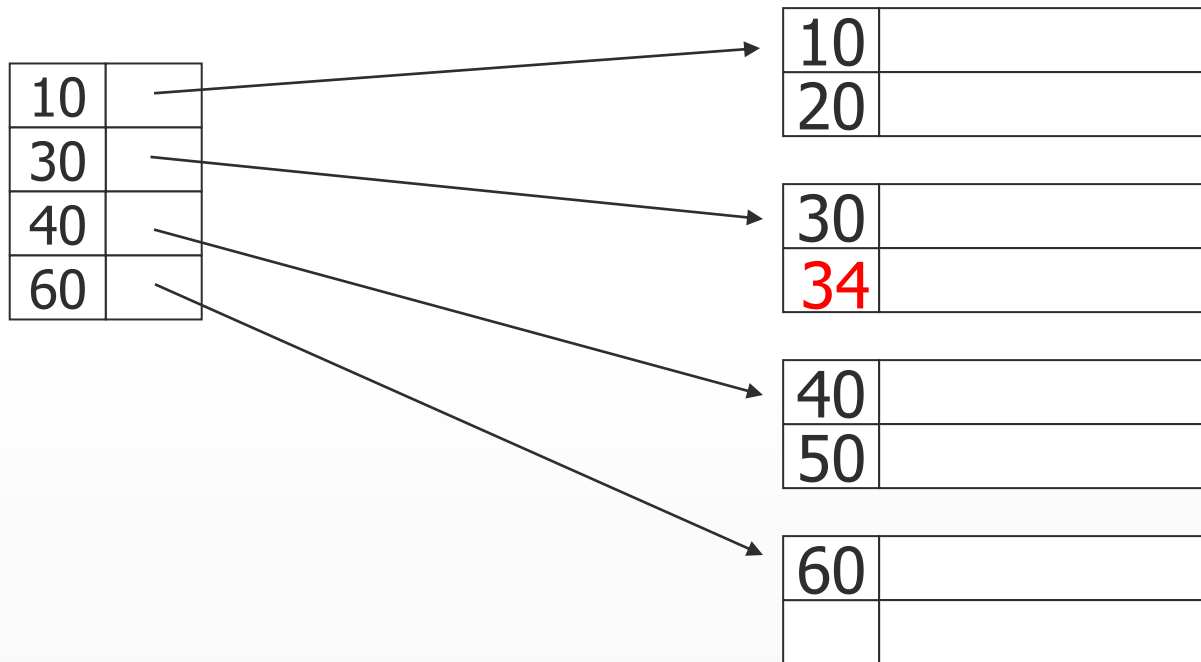


Dodavanje redak indeks



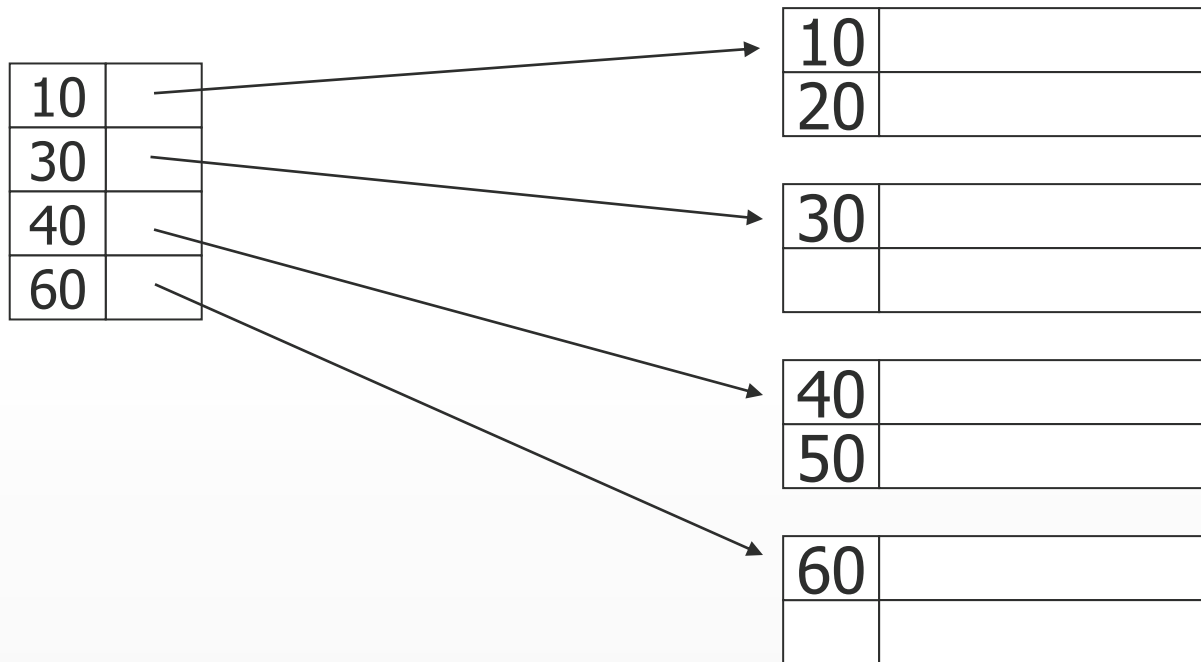
–insert record 34

Dodavanje redak indeks



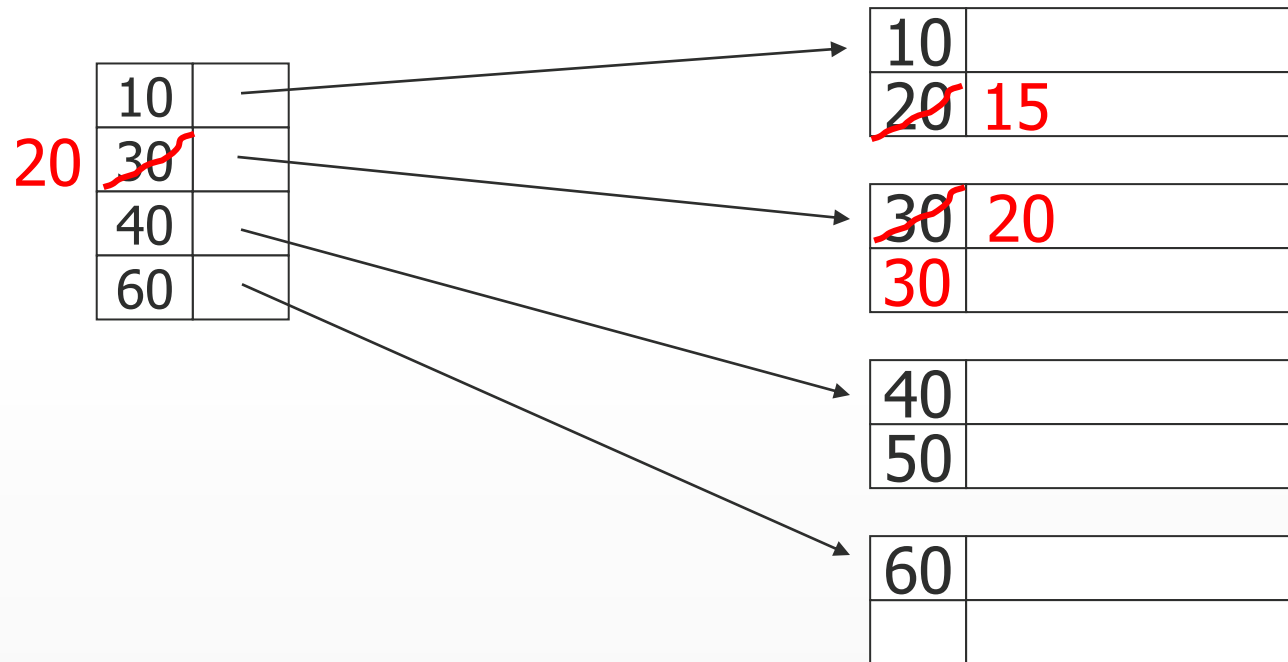
–insert record 34

Dodavanje redak indeks



–insert record 15

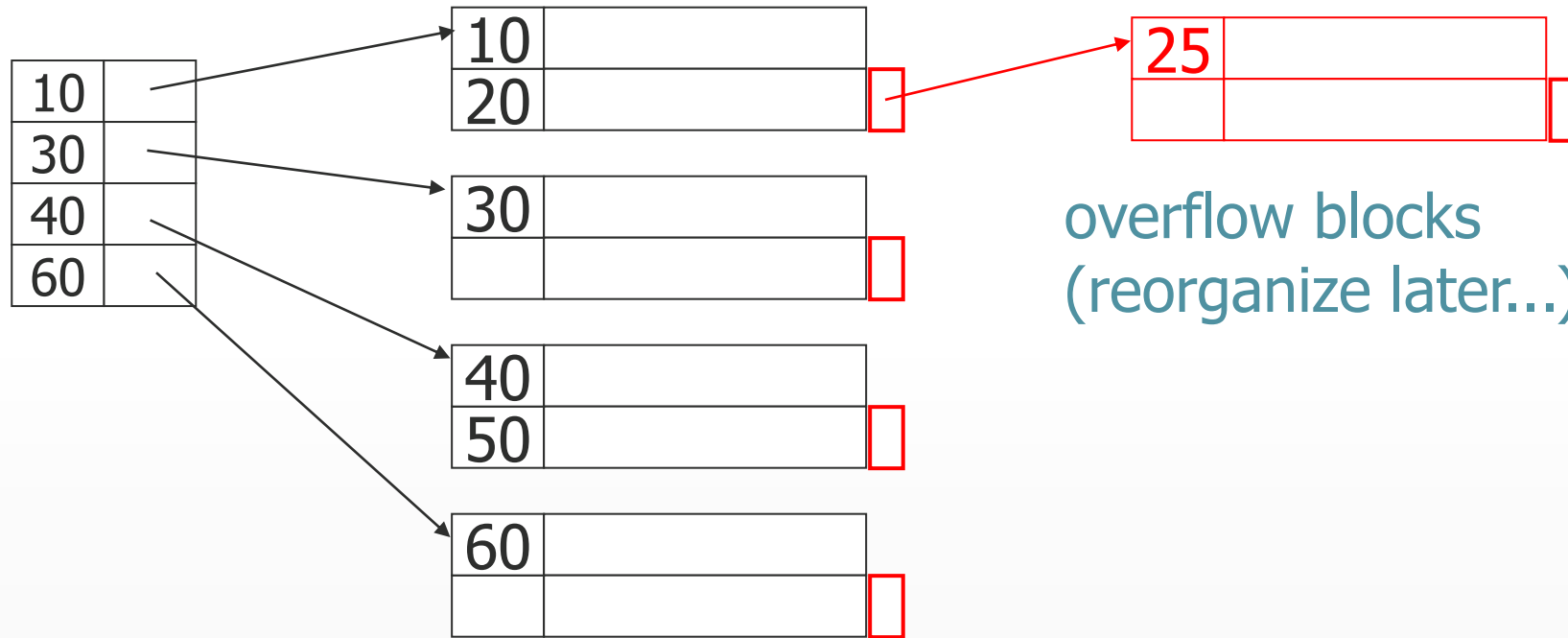
Dodavanje redak indeks



- Illustrated: Immediate reorganization
- Variation:
 - insert new block (chained file)
 - update index

–insert record 15

Dodavanje redak indeks



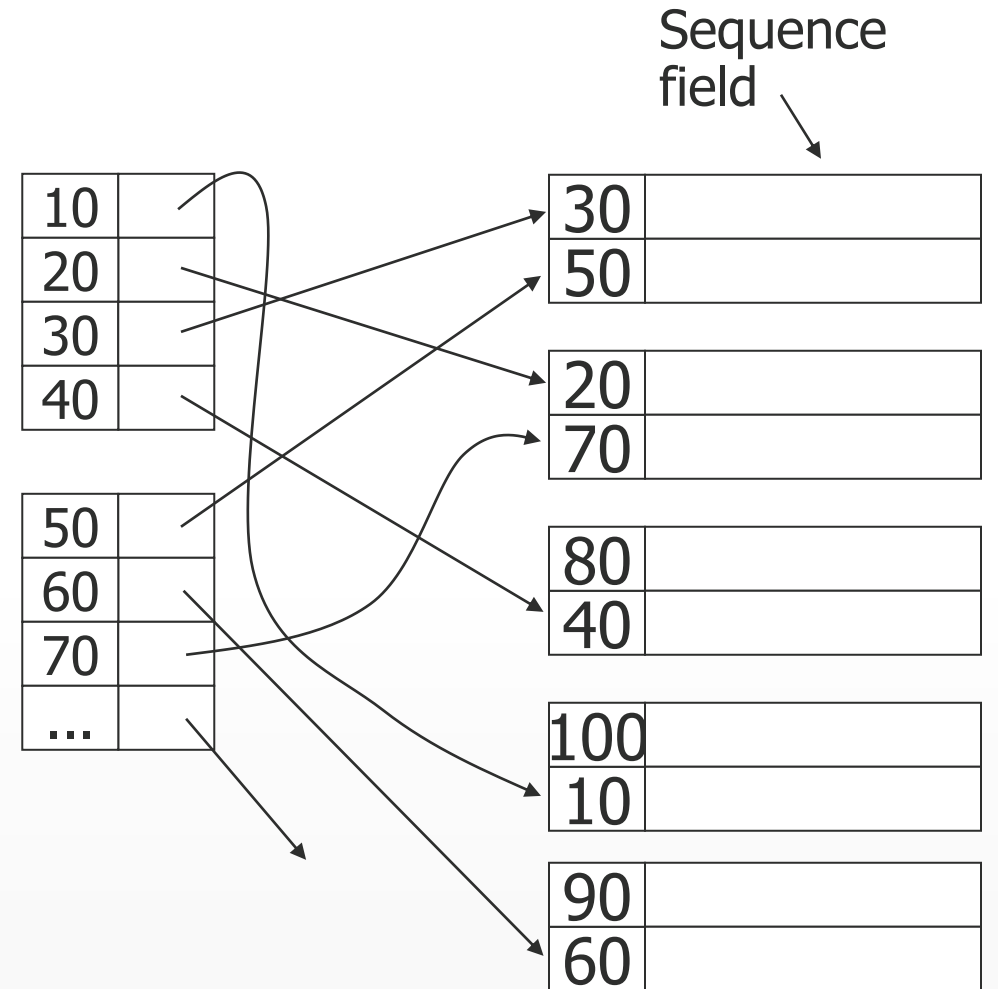
–insert record 25

Dodavanje gust indeks

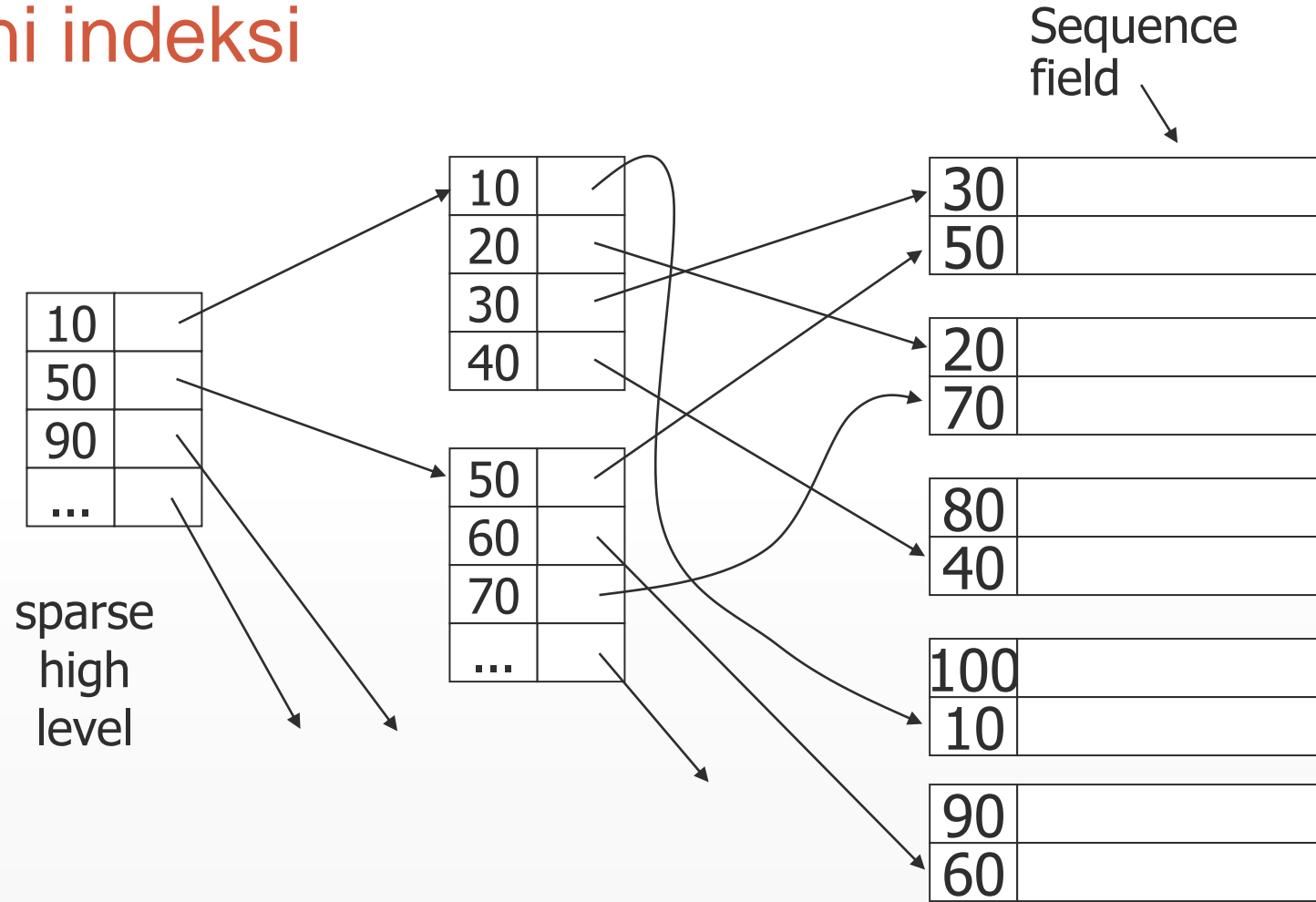
Analogno retkom

Sekundarni indeksi

- Indeksirano polje ne odgovara kriterijumu uređenja fajla sa podacima.
- Indeks mora biti **gust**.
- Pointer u slogovima indeksa mora biti **pointer ka slogu podataka**.

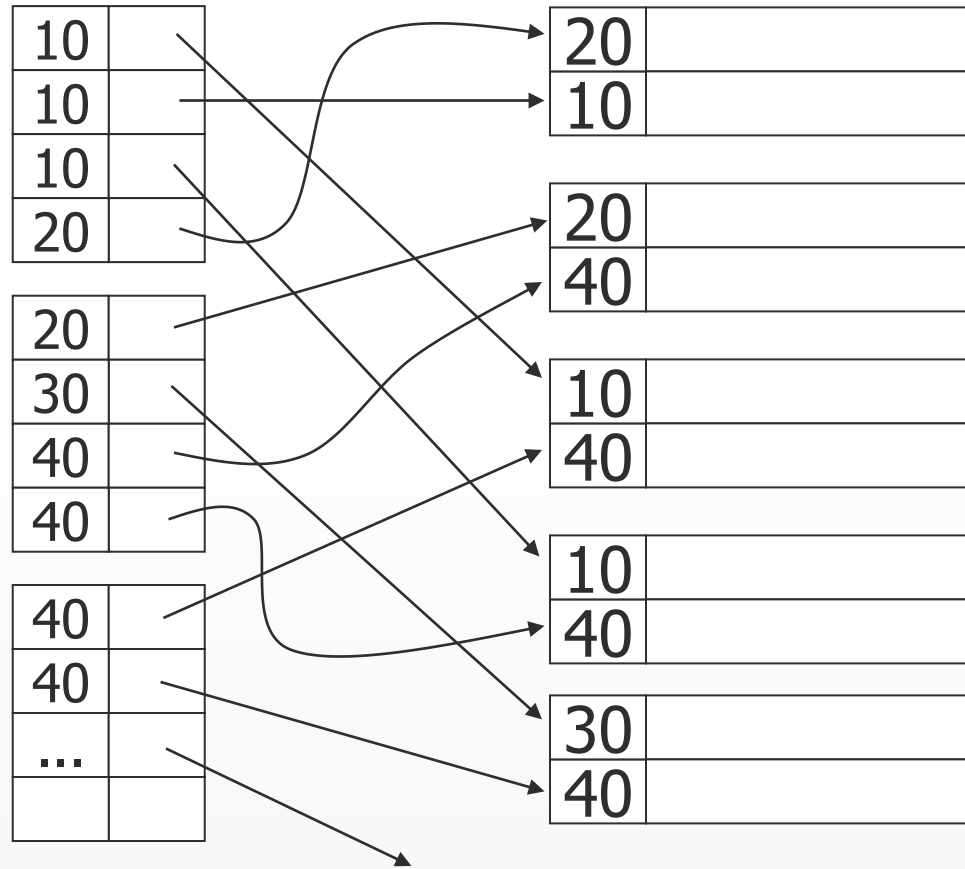


Sekundarni indeks



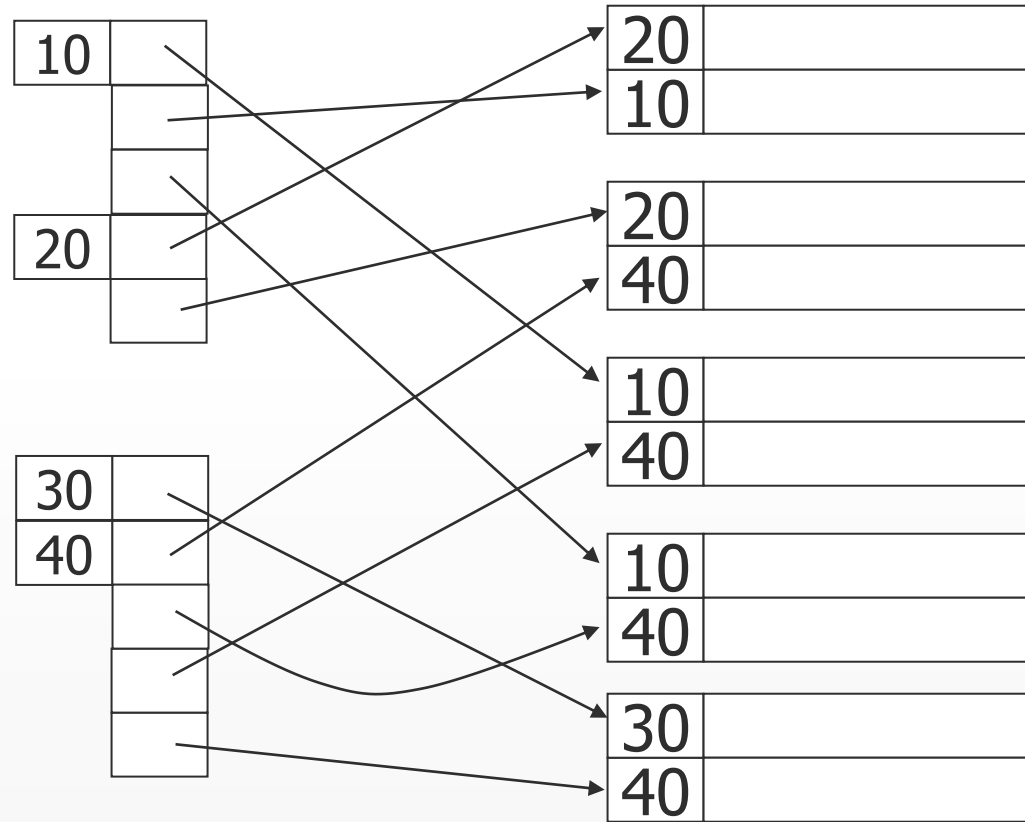
Duplikati i sekundarni indeksi

Problem:
excess overhead!
• disk space
• search time

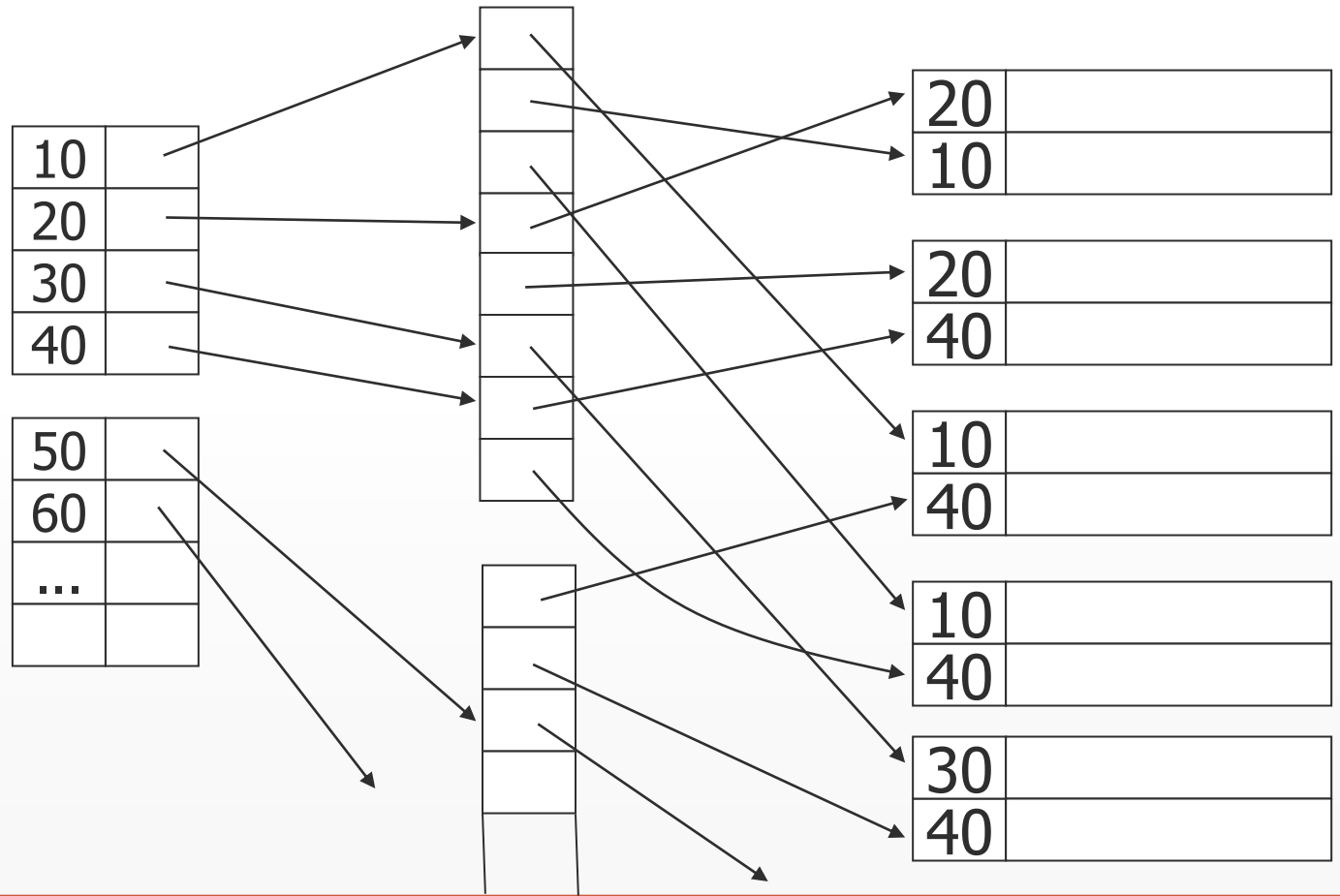


Duplikati i sekundarni indeksi

Problem:
variable size
records in
index!



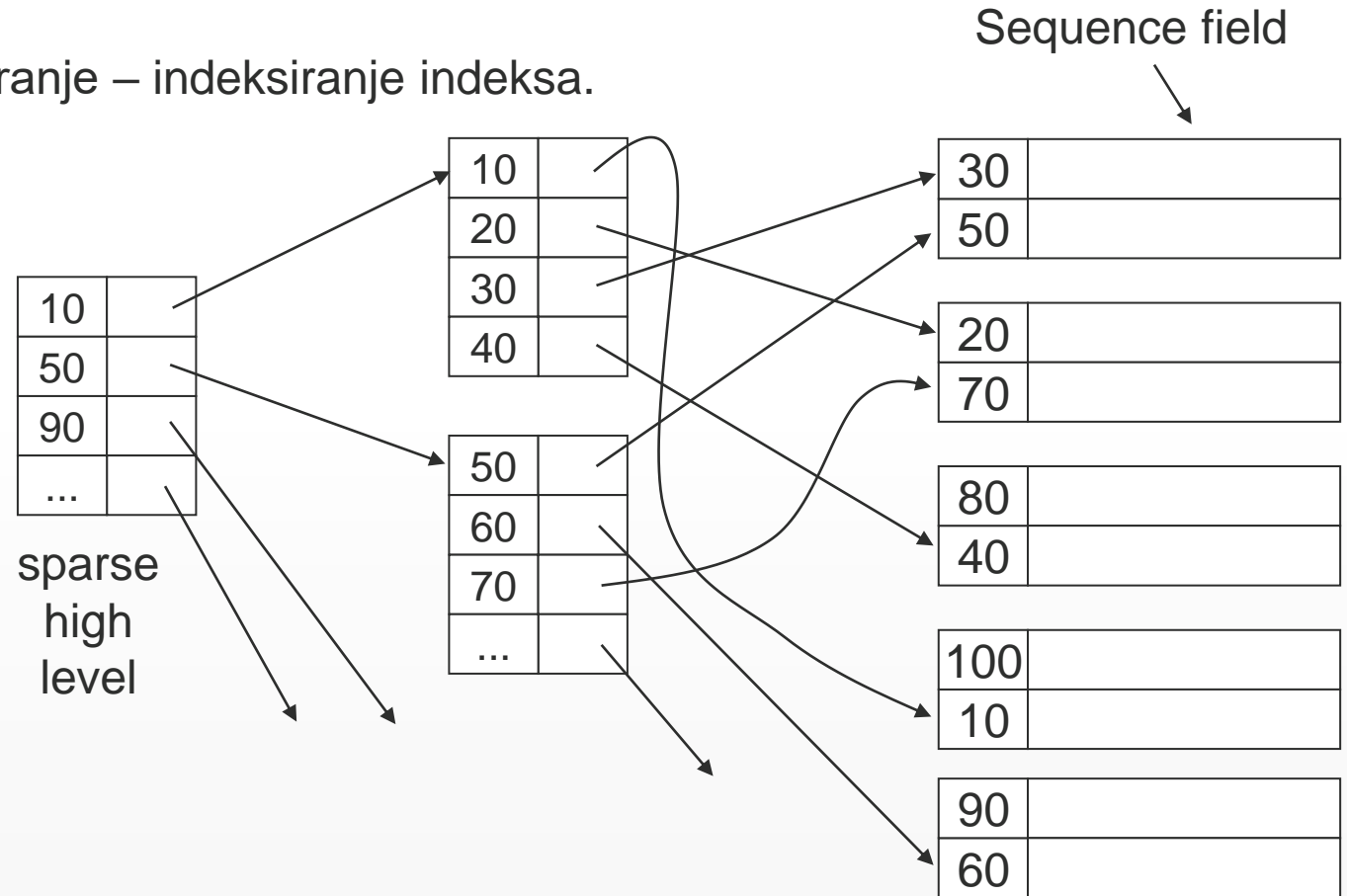
Duplikati i sekundarni indeksi



buckets

Višenivojski indeksi

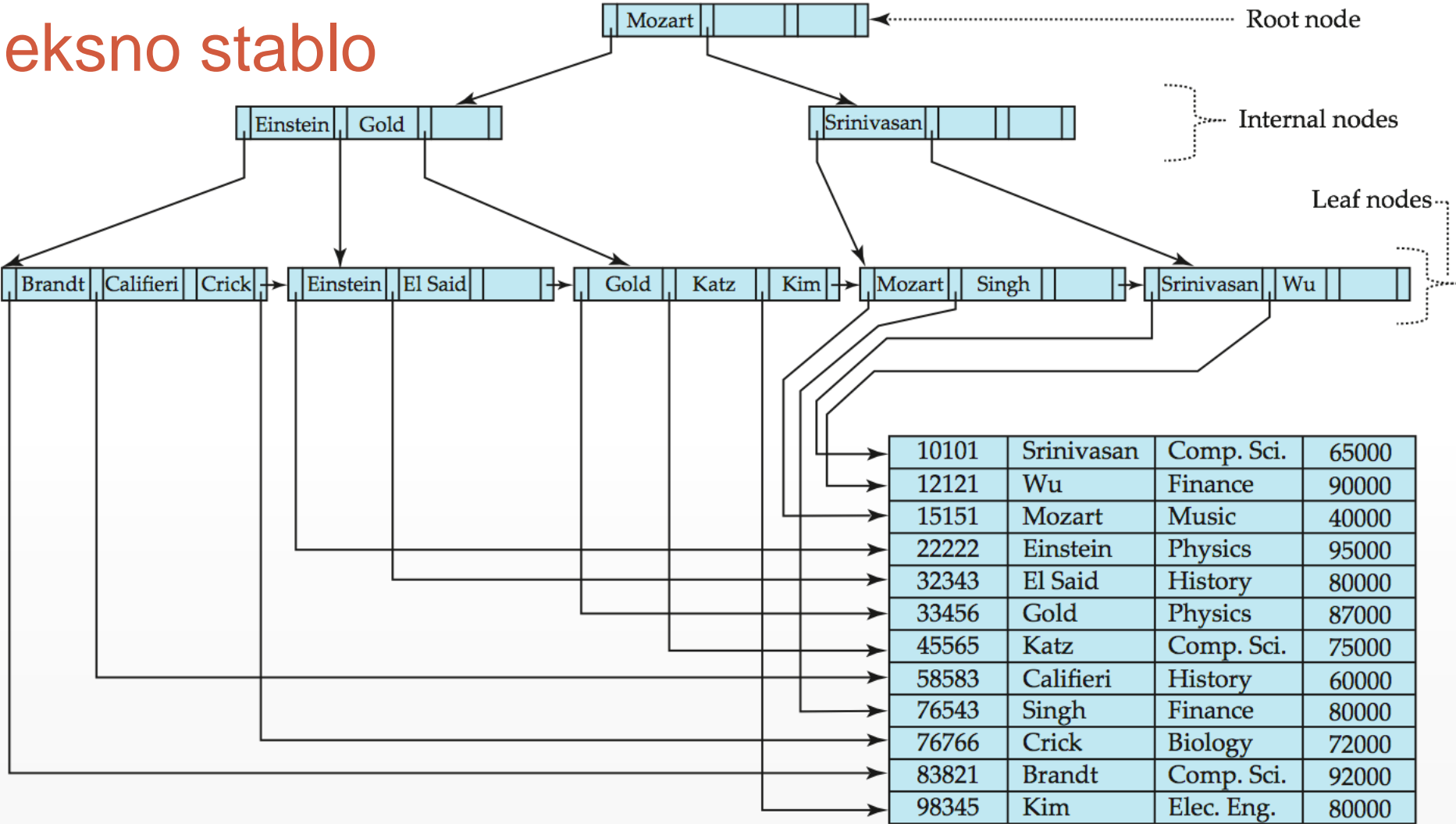
- Višenivojsko, hijerarhijsko indeksiranje – indeksiranje indeksa.



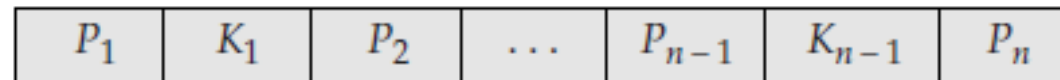
B+ indeksno stablo

- **B stablo** - balansirano binarno stablo; u njemu se svaka vrednost ključa pojavljuje jednom u stablu na nekom njegovom nivou zajedno sa pokazivačem na ostale podatke odgovarajućeg sloga ili bloka.
 - **B+ stablo** – B stablo u kome se pokazivači na ostale podatke nalaze samo u listovima stabla.
 - Svi listovi B+ stabla imaju isti nivo i međusobno su spregnuti pokazivačima.
 - Balansirana struktura.
 - Pristup, dodavanje i brisanje - $O(\log(n))$
 - Optimizovan za čitanje/pisanje velikih blokova podataka
-

B+ indeksno stablo

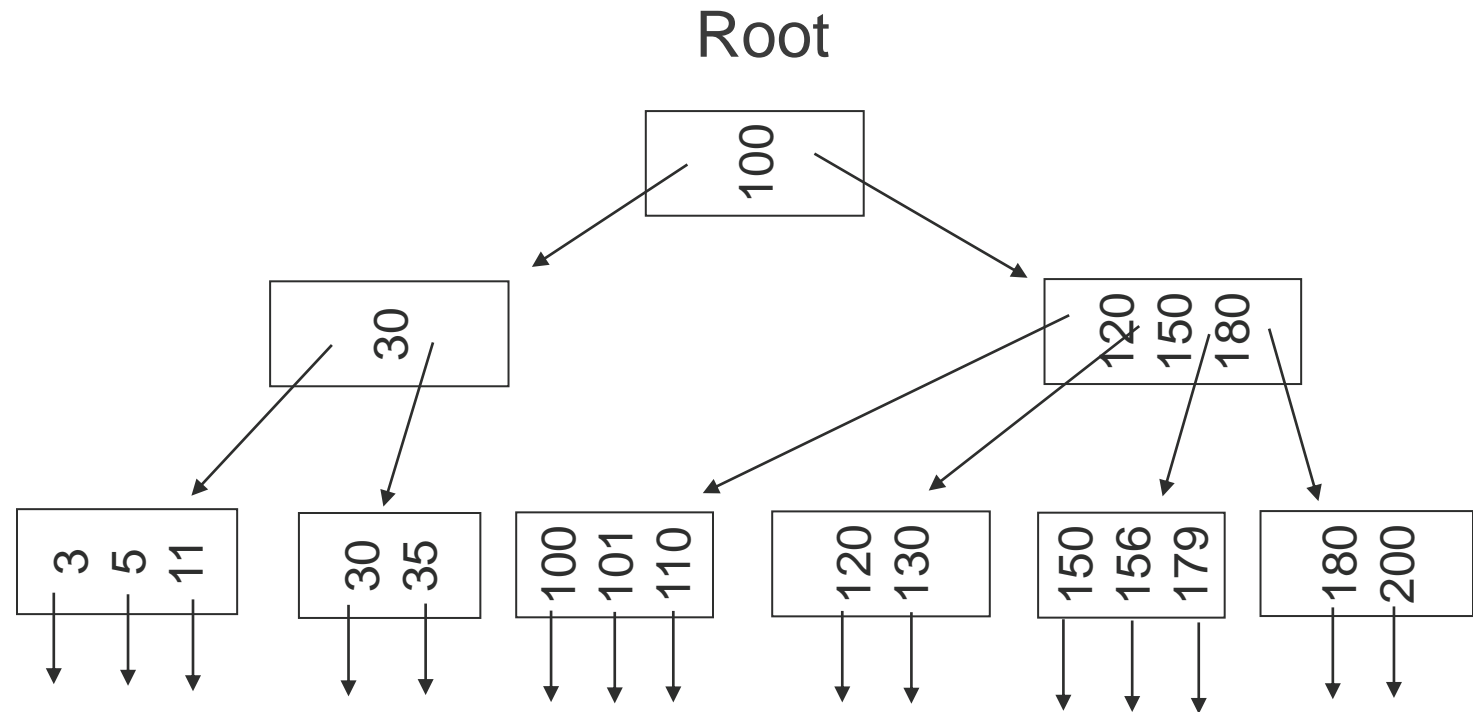


B+ indeksno stablo

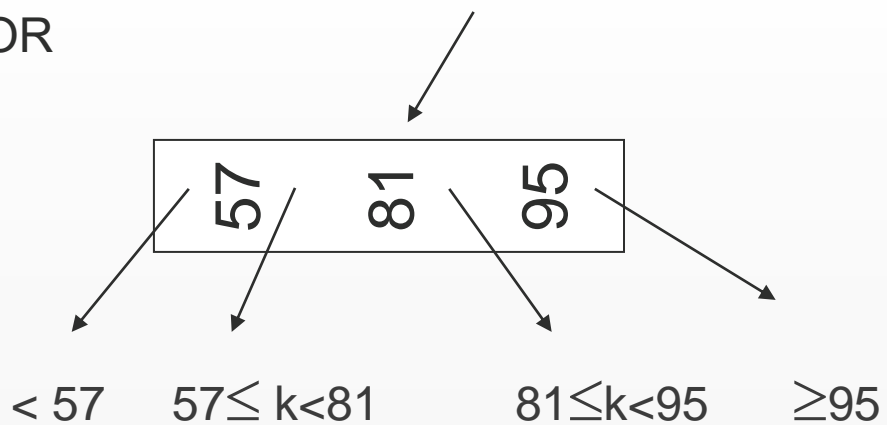


- K_i su vrednosti indeksiranog polja
 P_i pokazivači na decu (ako nisu listovi u pitanju) ili pointeri na slogove u fajlu sa podacima.
 - B+ je balansirano stablo u kojem svaki put pretrage od korena do lista ima istu dužinu
 - Svaki unutrašnji čvor je najmanje do pola pun ($N/2 - 1 \leq \text{num of keys} \leq N - 1$).
 - Svaki unutrašnji čvor sa k ključeva ima $k+1$ nenul dete.
 - Listovi sadrže:
 - Pokazivače na lokacije torke
 - Kompletne torke (ređi slučaj).
-

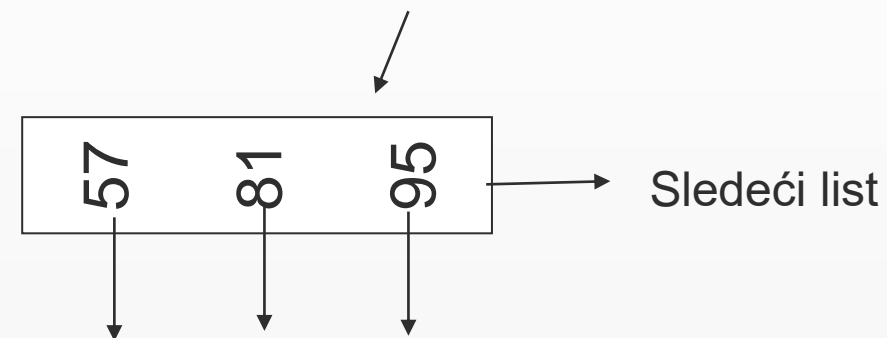
B+ indeksno stablo



ČVOR



LIST



B+ indeksno stablo

Dodavanje / brisanje

[B+ Tree Visualization \(usfca.edu\)](http://usfca.edu)

Indeksiranje vs Heširanje

Organizacija fajlova

Poređenje

- Ključni faktor poređenja jeste vrsta upita.
- Upiti jednakosti ključa pretraživanja sa nekom konstanom i upiti po oblasti vrednosti ključa indeksiranja i heširanja.

```
SELECT ...  
FROM R  
WHERE R.A = 5
```

Broj pristupa pri indeksiranju je srazmeran logaritmu broja vrednosti za A, dok je pri **heširanju** nezavisan od veličine same relacije.

```
SELECT  
FROM R  
WHERE R.A > 5
```

Ovakav upit bi zahtevao heširanje funkcijom koja zadržava redosled ključeva u rasutoj datoteci. **Indeksiranje** je u ovom slučaju bolji izbor.

Indeksi u SQL-u

SQL

- Create index name on rel (attr)
 - Create unique index name on rel (attr)
 - Ne može se specificirati vrsta indeksa (B-stablo, Heš, ...) ili bilo kakvi parametri indeksa.
-

Višedimenzioni indeksi

Višedimenzioni indeksi

- **Višedimenziono indeksiranje** – indeksi nad više ključeva pretrage.
 - Posebno interesantno kod GIS-ma i u Data warehouse-u.
 - Koriste se za ubrzavanje pretrage po više ključeva.
 - Vrste:
 - Heš bazirani
 - Stabla
 - Bitmap indeksi
-

Primene

- **Višedimenziono indeksiranje** – indeksi nad više ključeva pretrage.
- Posebno interesantno kod GIS-ma i u Data v
- Koriste se za ubrzavanje pretrage po više klj

Applications: GIS

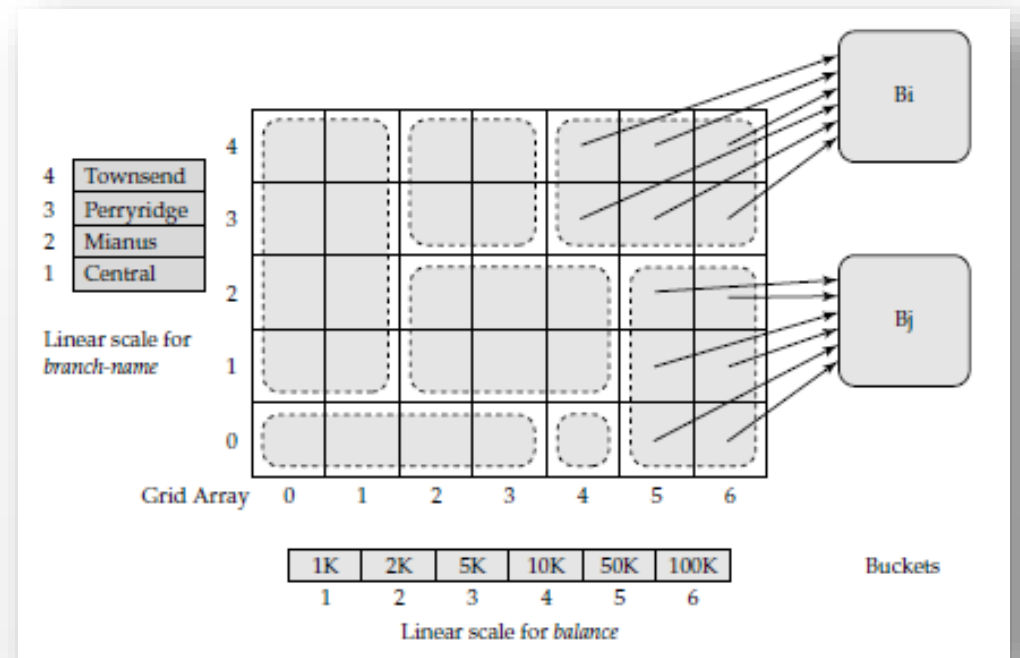
- Geographic Information Systems
 - Objects are in two dimensional space
 - The objects may be points or shapes
 - Objects may be houses, roads, bridges, pipelines and many other physical objects.
- Query types
 - Partial match queries
 - Specify the values for one or more dimensions and look for all points matching those values in those dimensions.
 - Range queries
 - Set of shapes within the range
 - Nearest neighbor queries
 - Closest point to a given point
 - Where-am-I queries
 - When you click a mouse, the system determines which of the

Data Cubes

- Data exists in high dimensional space
- A chain store may record each sale made, including
 - The day and time
 - The store in which the sale was made
 - The item purchased
 - The color of the item
 - The size of the item
- Attributes are seen as dimensions multidimensional space, data cube.
- Typical query
 - Give a class of pink shirts for each store and each month of 1998.

Struktura rešetke

- Za upite po n ključeva – n-dimenziona rešetka.
- Svaka dimenzija odgovara jednom ključu pretraživanja.



Heš bazirani više dimenzioni indeksi rešetke

- Primer

Salary

0-20K	1
20K-50K	2
50K- ∞	3

Grid

1	2	3
Toy	Sales	Personnel



EX:

h1(toy)	=0	000	
h1(sales)	=1	001	
h1(art)	=1	010	
:		011	
h2(10k)	=01	100	
h2(20k)	=11	101	
h2(30k)	=01	110	
h2(40k)	=00	111	
:			



<Fred,toy,10k>, <Joe,sales,10k>
<Sally,art,30k>

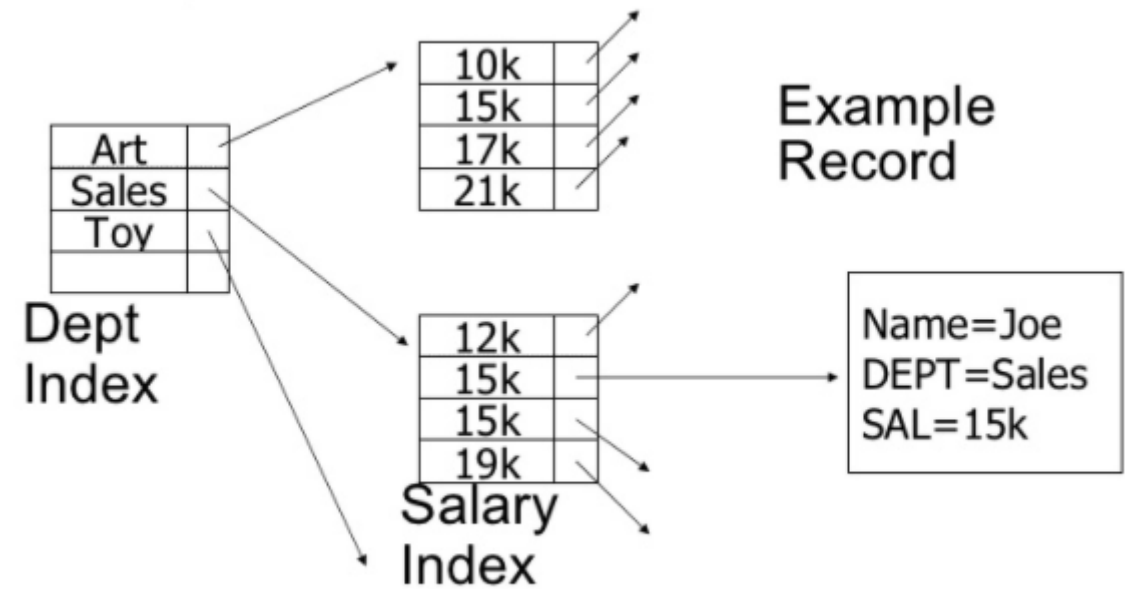
Višedimenzioni indeksi tipa stabla

- Vrste:
 - Indeksi sa višestrukim ključevima
 - Kd-stabla
 - Kvadratna stabla
 - R-stabla
-

Indeksi sa višestrukim ključevima

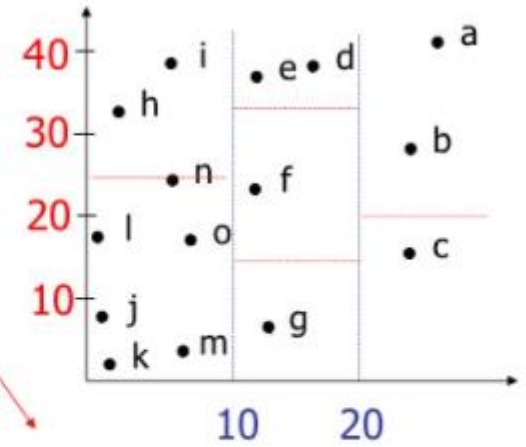
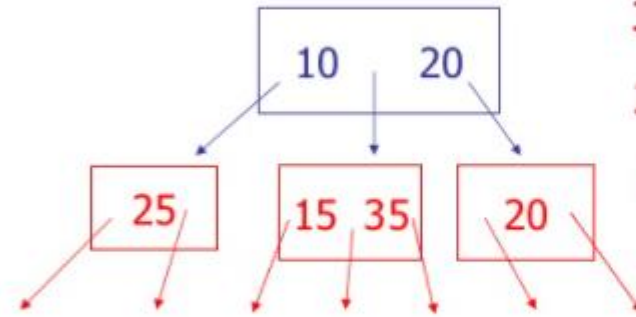
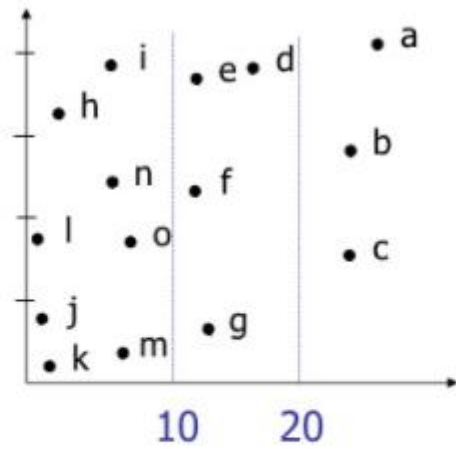
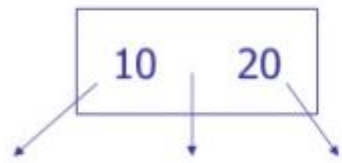
- Više atributa koji predstavljaju dimenzije po kojima se razvrstavaju slogovi/podaci.
- Indeks nad indeksom u kom su na istom nivou indeksi nad jednim atributom

Example

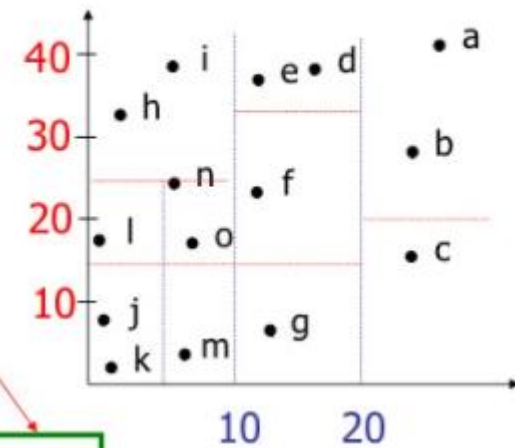
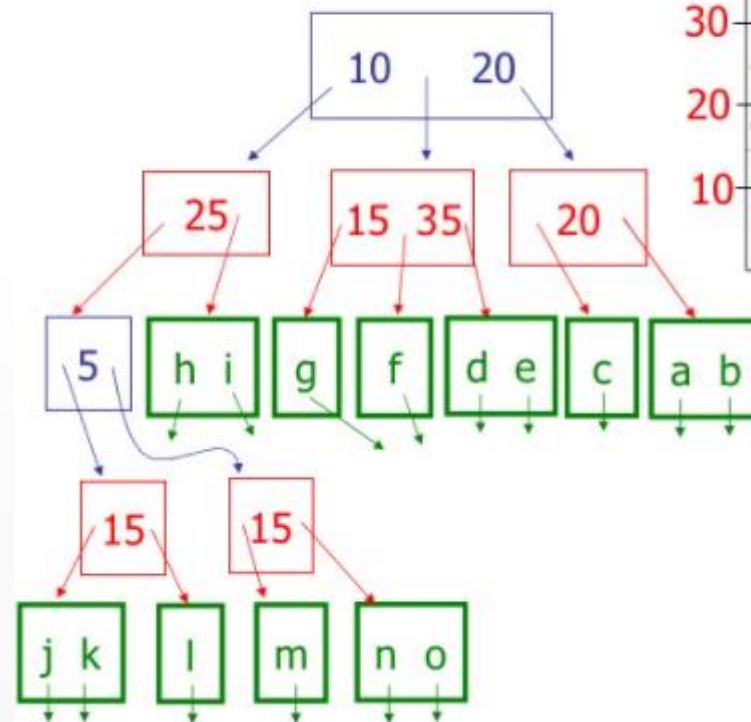
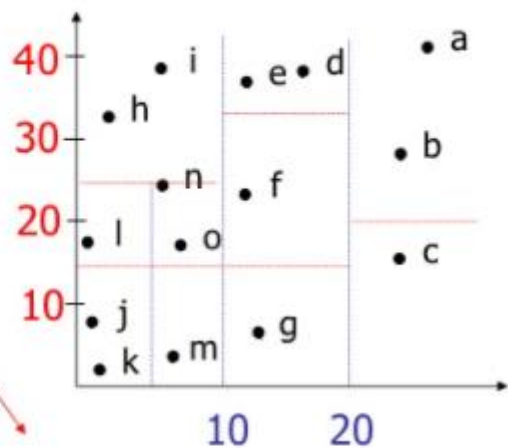
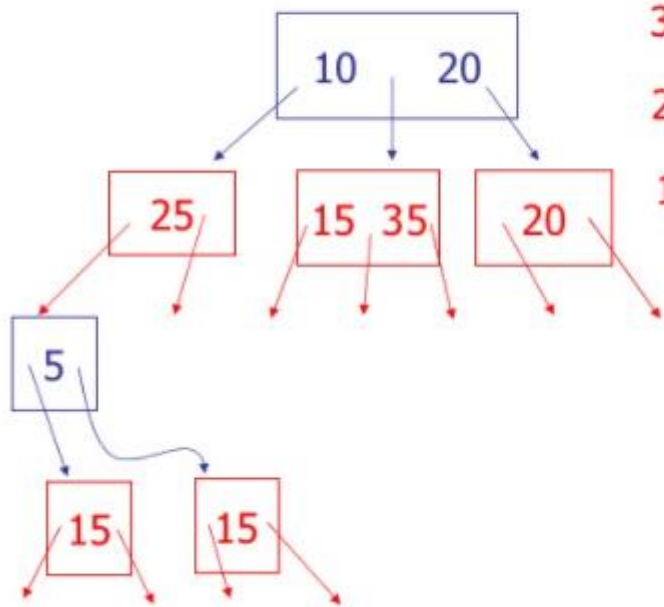


KD stabla

- K-dimenziono stablo.



Višedimenzioni indeksi tipa stabla



Bitmap indeksi

- Bitmap indeks po ključu pretrage je binarni niz čija je dužina jednaka broju slogova u relaciji.
- Za svaki ključ pretrage se formira jedan bitmap indeks.
- Za upite po više ključeva se primenjuju bitovske operacije
presek - AND
unija – OR
komplement - NOT

Males with income level L1:
 $10010 \text{ AND } 10100 = 10000$

record number	ID	gender	income_level
0	76766	m	L1
1	22222	f	L2
2	12121	f	L1
3	15151	m	L4
4	58583	f	L3

Bitmaps for *gender*

m	10010
f	01101

Bitmaps for *income_level*

L1	10100
L2	01000
L3	00001
L4	00010
L5	00000

Bitmap indeksi

- Brisanje

Može se regulisati **bitmapom egzistencije**

$\text{not}(A=v):$ (NOT bitmap-A-v) AND ExistenceBitmap

- Null vrednosti se mogu rešiti na isti način
 - Za velike relacije bitmapa je duga, ali se može kompresovati.
-