

Strukture podataka

Indeksne strukture

Ispitni zadaci

II kolokvijum - 30.05.2016.

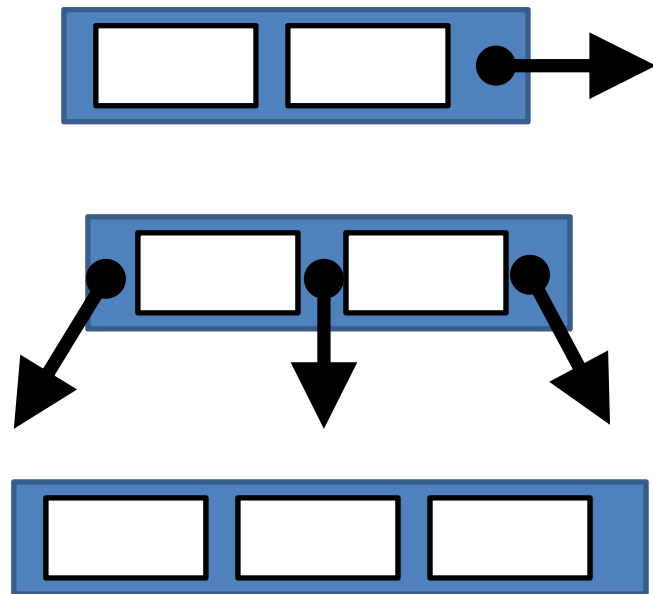
(15 poena) Opšta stabla i indeksne strukture

- (5 poena) Za B+ stablo reda 3 prikazati kako izgleda postupak formiranja stabla nakon dodavanja sledećih ključeva: 1, 2, 16, 6, 15, 7, 4, 5, 8, 10, 9. Prilikom cepanja čvora smatrati da veći broj ključeva ide u levog suseda.
- (10 poena) Baza podataka sadrži 3 000 000 slogova dužine 250 B. Smatrati da slogovi nisu uređeni ni po jednom kriterijumu. Veličina bloka na disku je 4096 B, veličina ključa 15 B, a veličina pokazivača na blok je 6 B. Odrediti:
 - koliko MB na disku zauzima osnovna datoteka, (3 poena),
 - ako se za indeksiranje koristi B stablo, odrediti koliko puta se ubrzava pristup i za koliko procenata se poveća datoteka (7 poena).

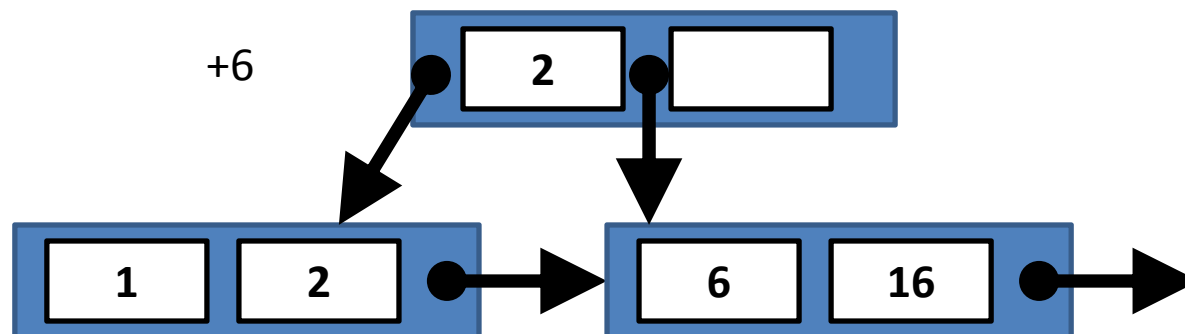
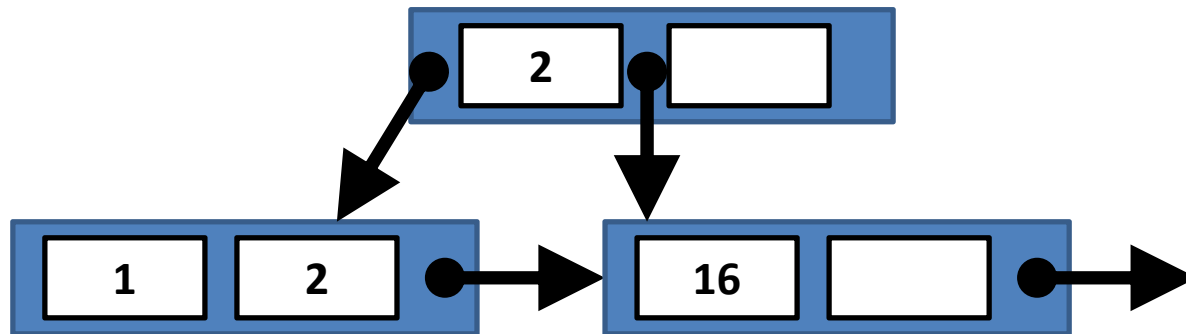
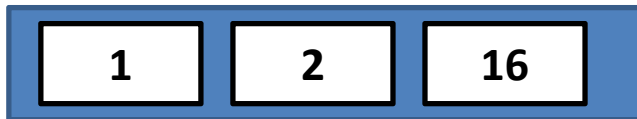
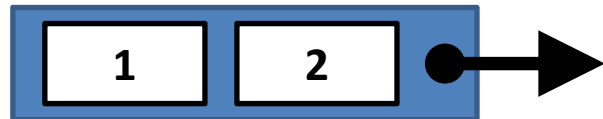
Za **B+ stablo reda 3** prikazati kako izgleda postupak formiranja stabla nakon dodavanja sledećih ključeva: 1, 2, 16, 6, 15, 7, 4, 5, 8, 10, 9. Prilikom cepanja čvora smatrati da veći broj ključeva ide u levog suseda.

B+ stablo – pri cepanju lista, zadnji ključ iz levog se kopira u roditelja, unutrašnji čvorovi se ponašaju kao B stablo.

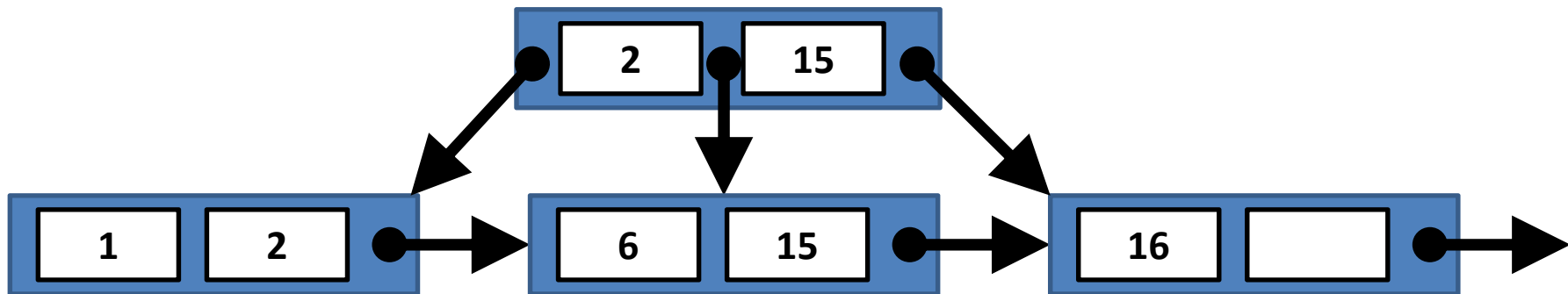
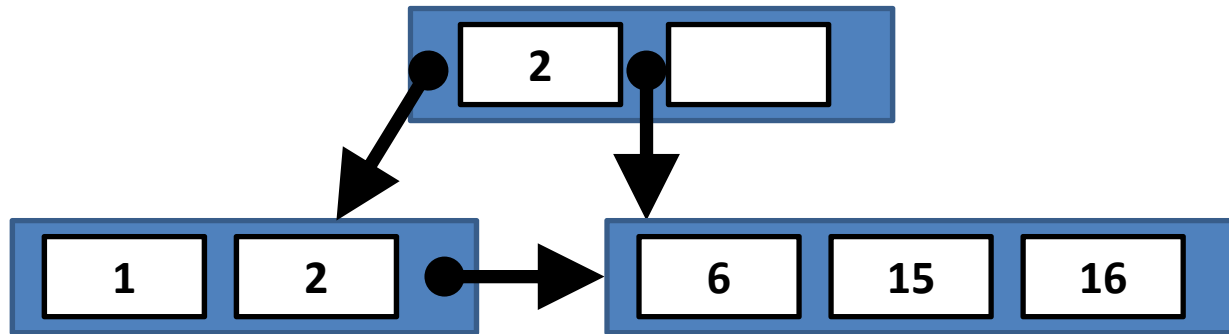
reda 3 – dva ključa po čvoru.



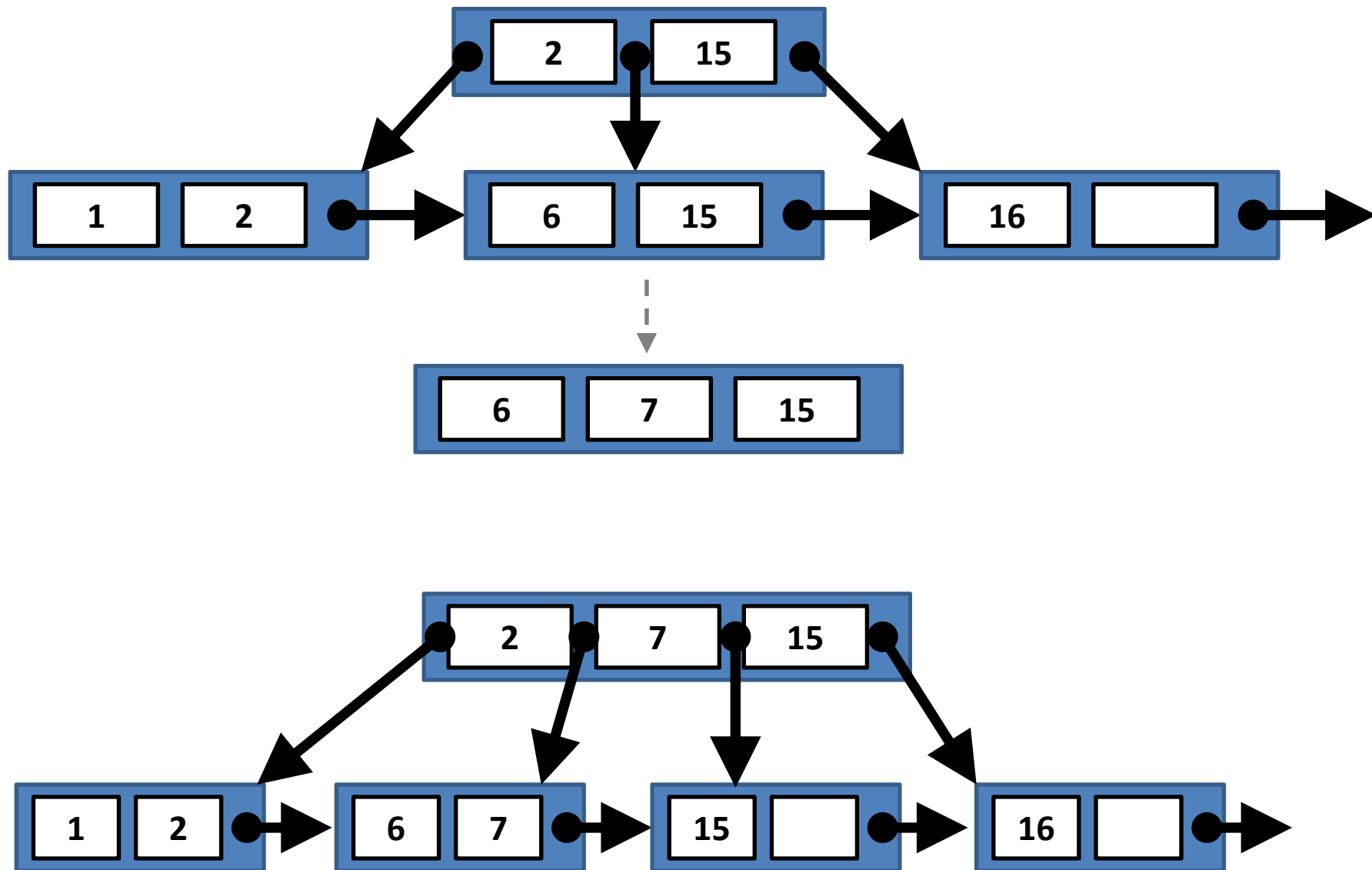
1, 2, 16, 6, 15, 7, 4, 5, 8, 10



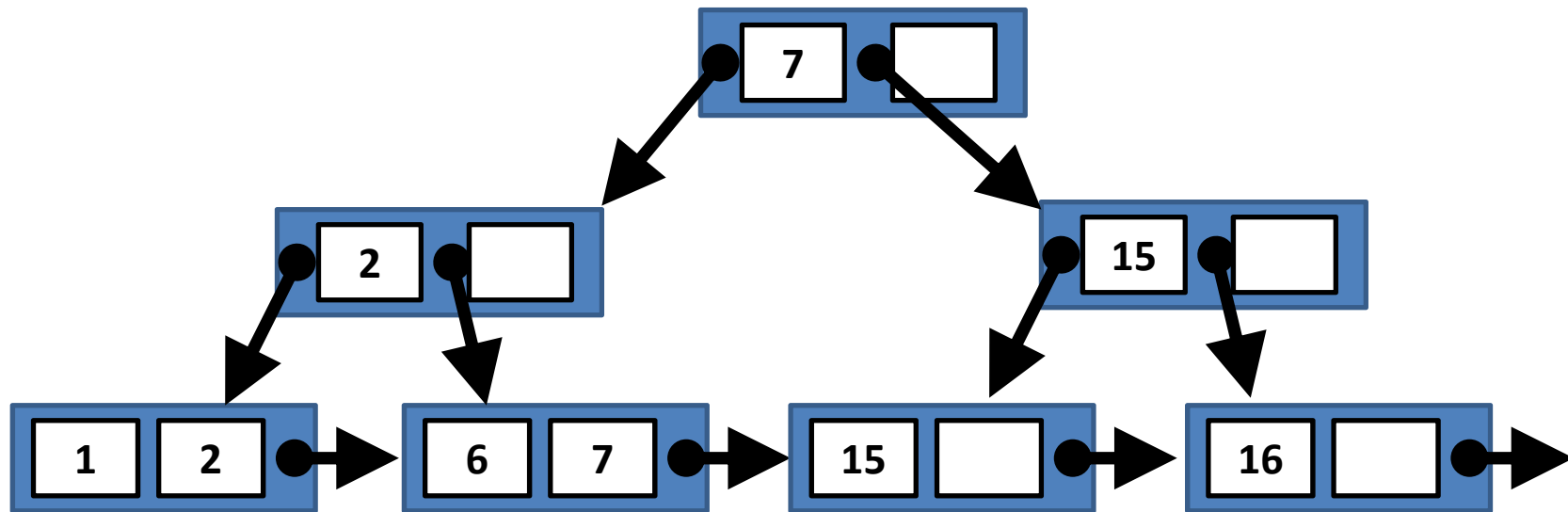
1, 2, 16, 6, **15**, 7, 4, 5, 8, 10



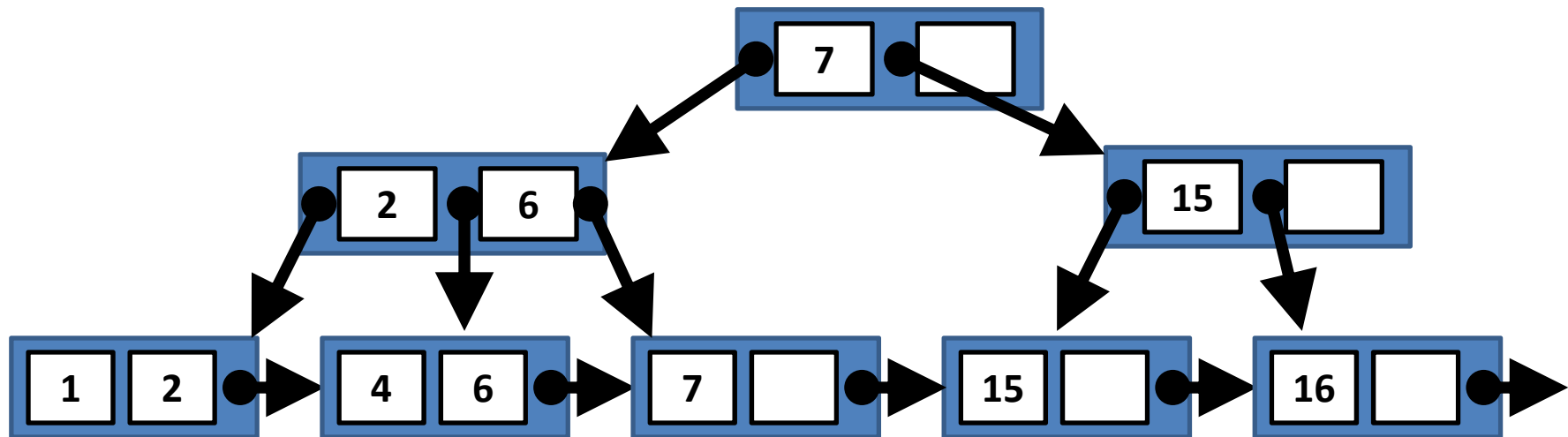
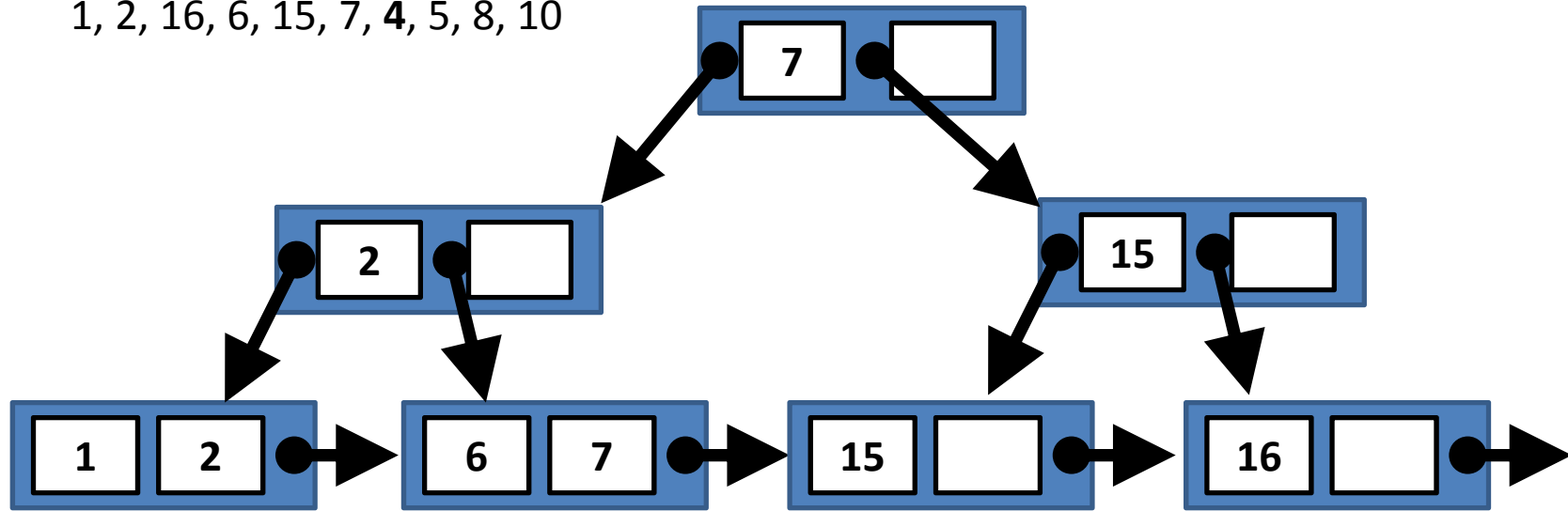
1, 2, 16, 6, 15, **7**, 4, 5, 8, 10



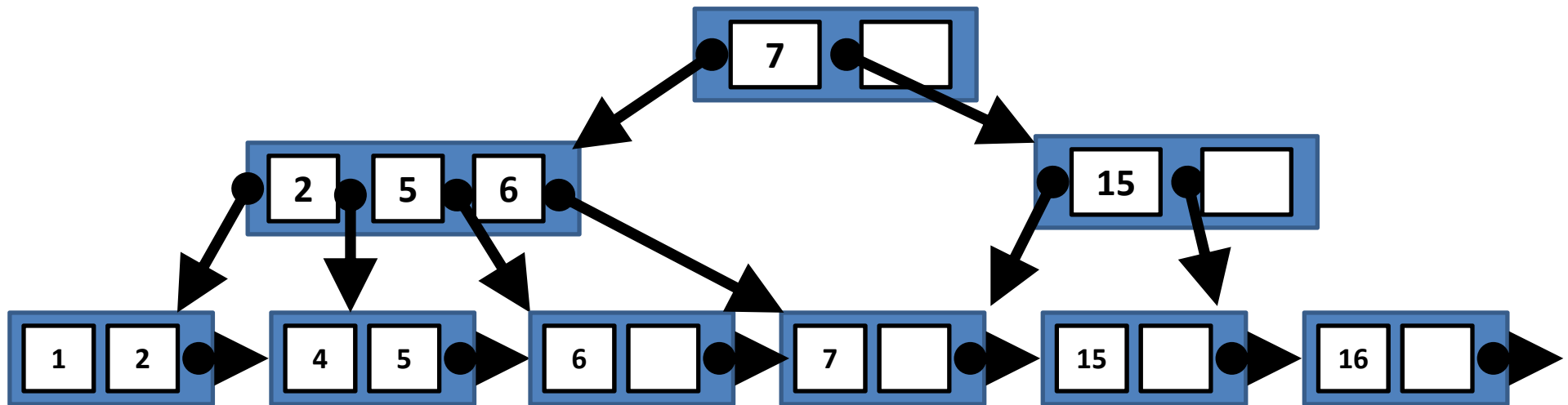
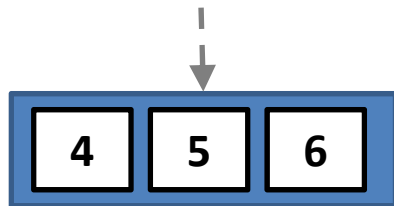
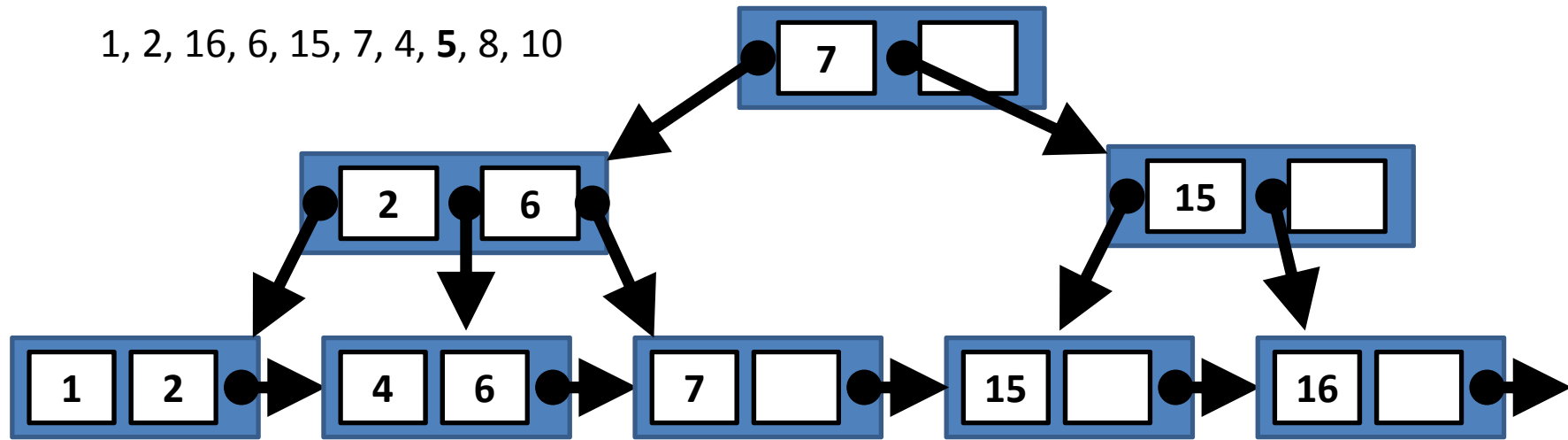
1, 2, 16, 6, 15, 7, 4, 5, 8, 10



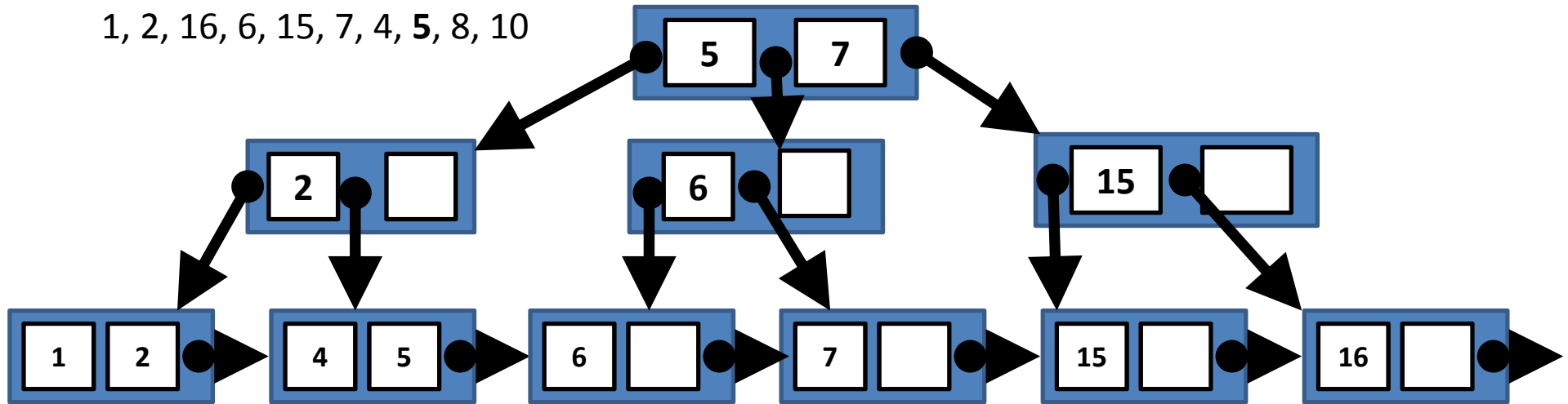
1, 2, 16, 6, 15, 7, 4, 5, 8, 10



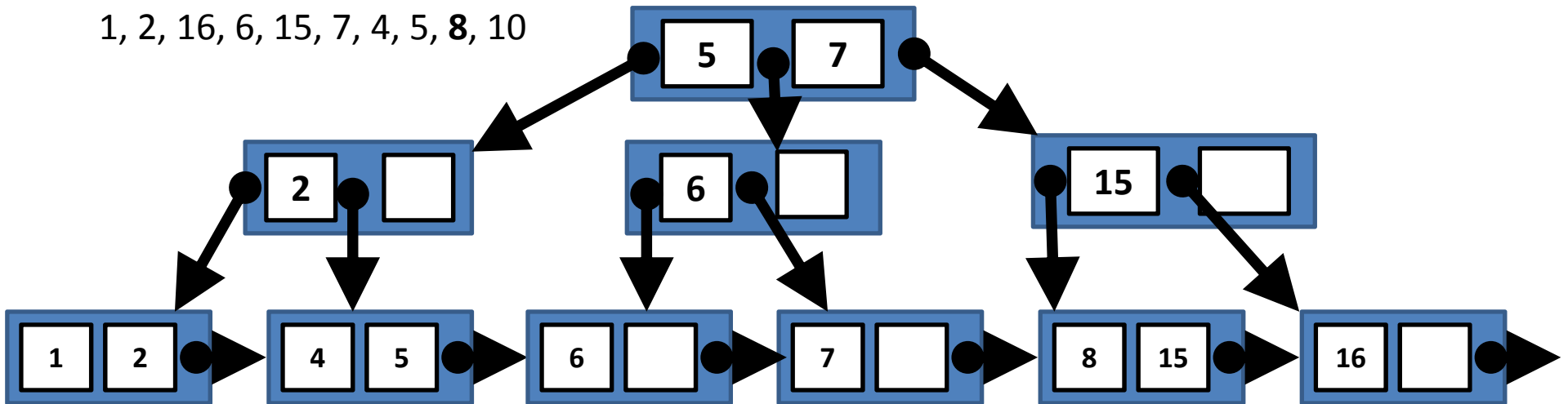
1, 2, 16, 6, 15, 7, 4, 5, 8, 10



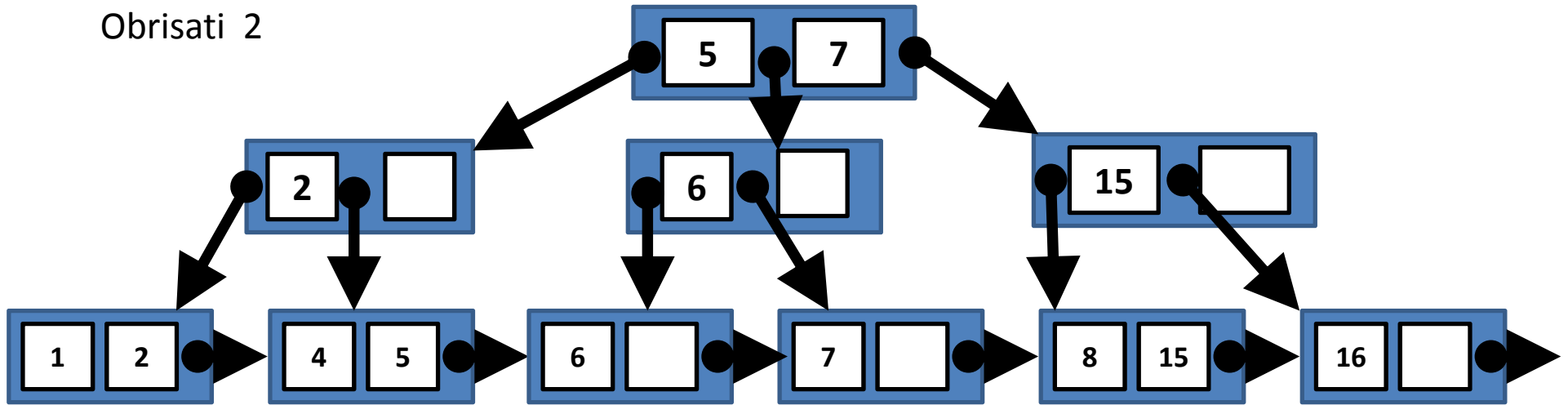
1, 2, 16, 6, 15, 7, 4, 5, 8, 10



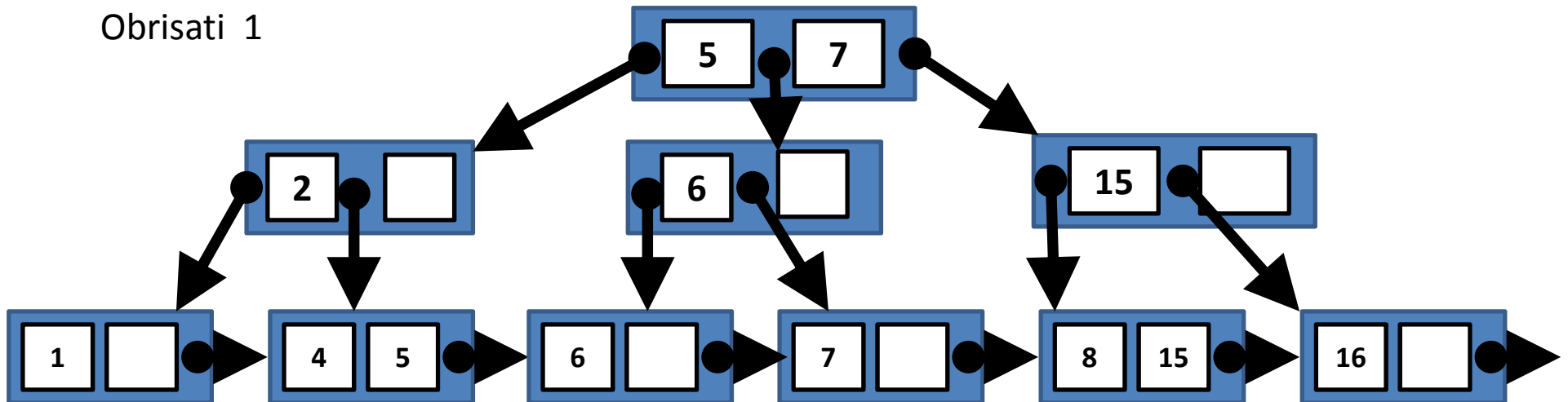
1, 2, 16, 6, 15, 7, 4, 5, 8, 10



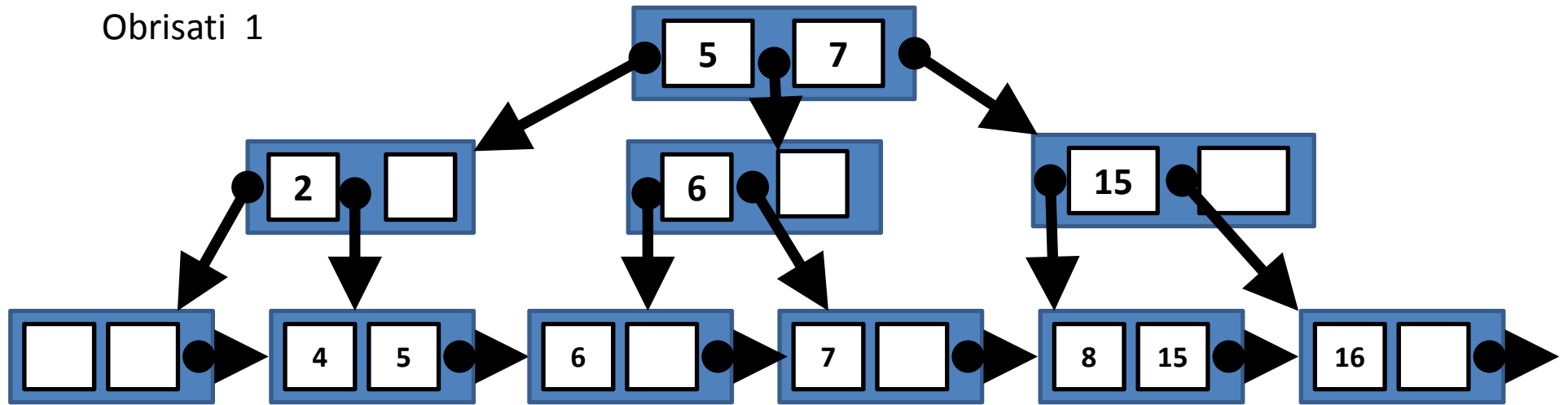
Obrisati 2



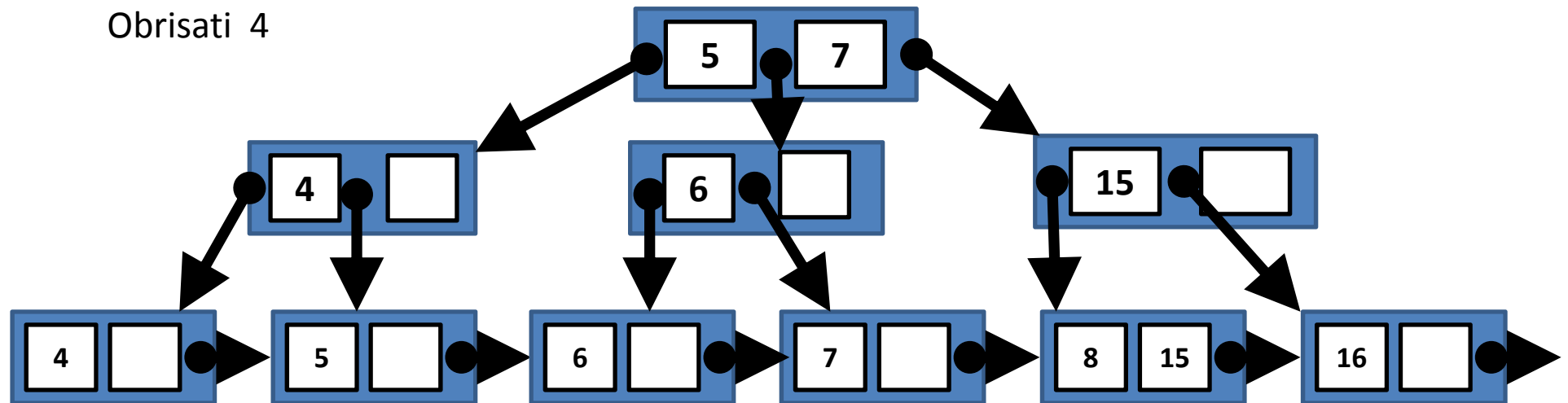
Obrisati 1



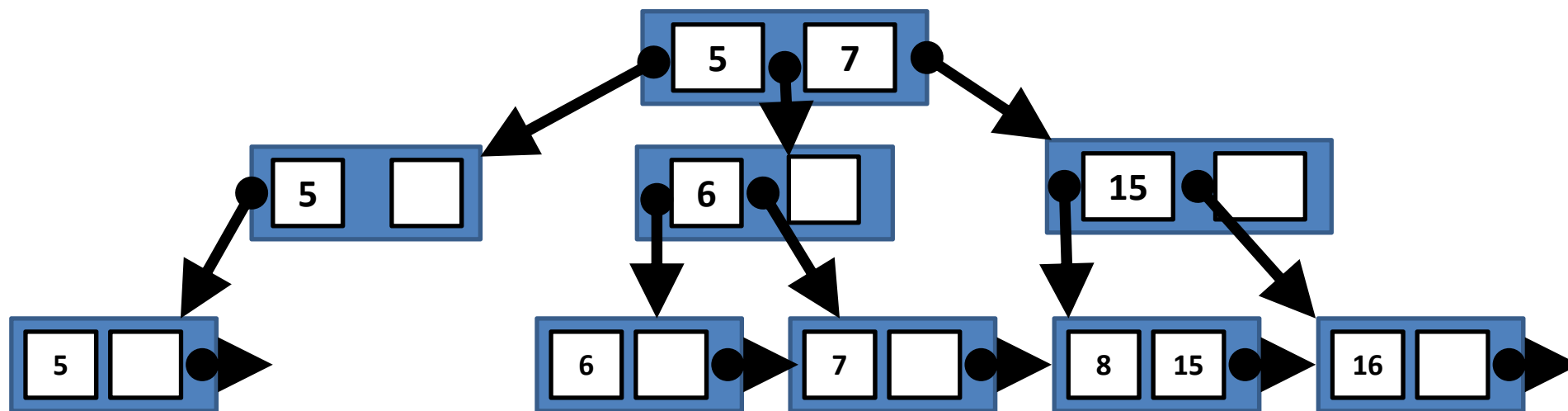
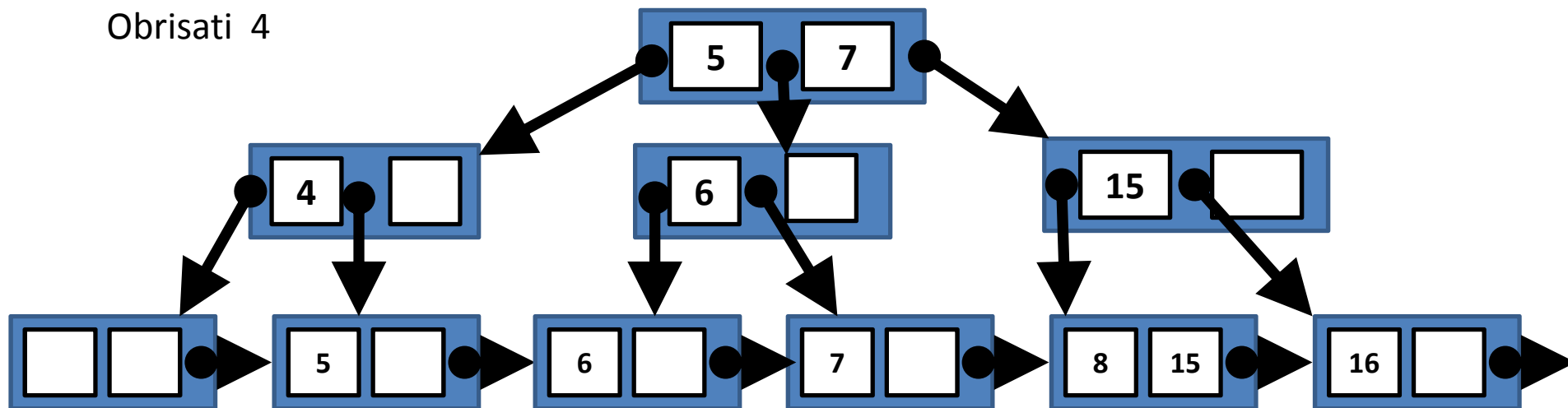
Obrisati 1

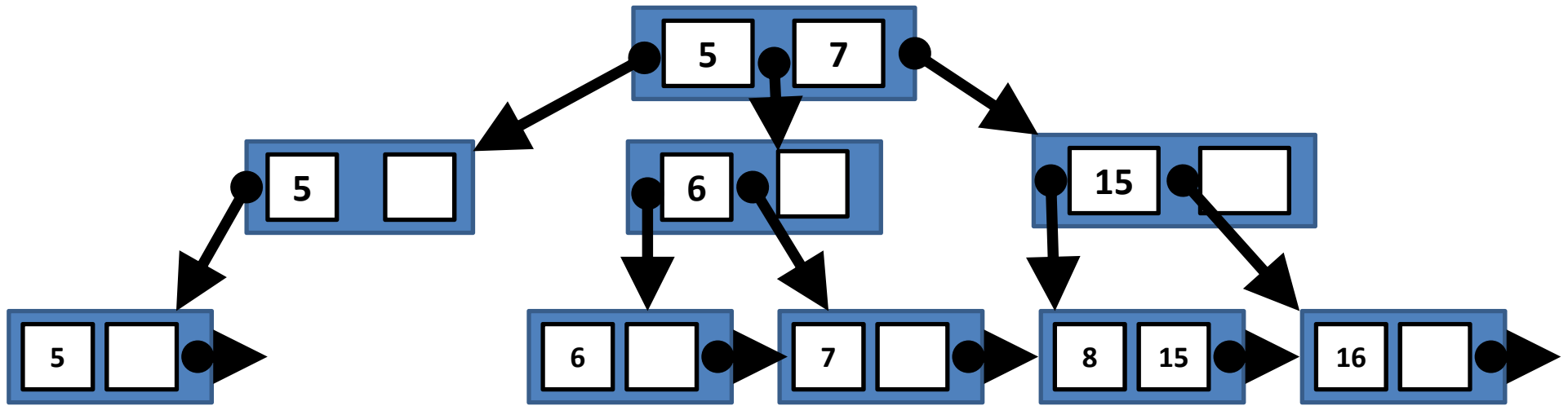


Obrisati 4

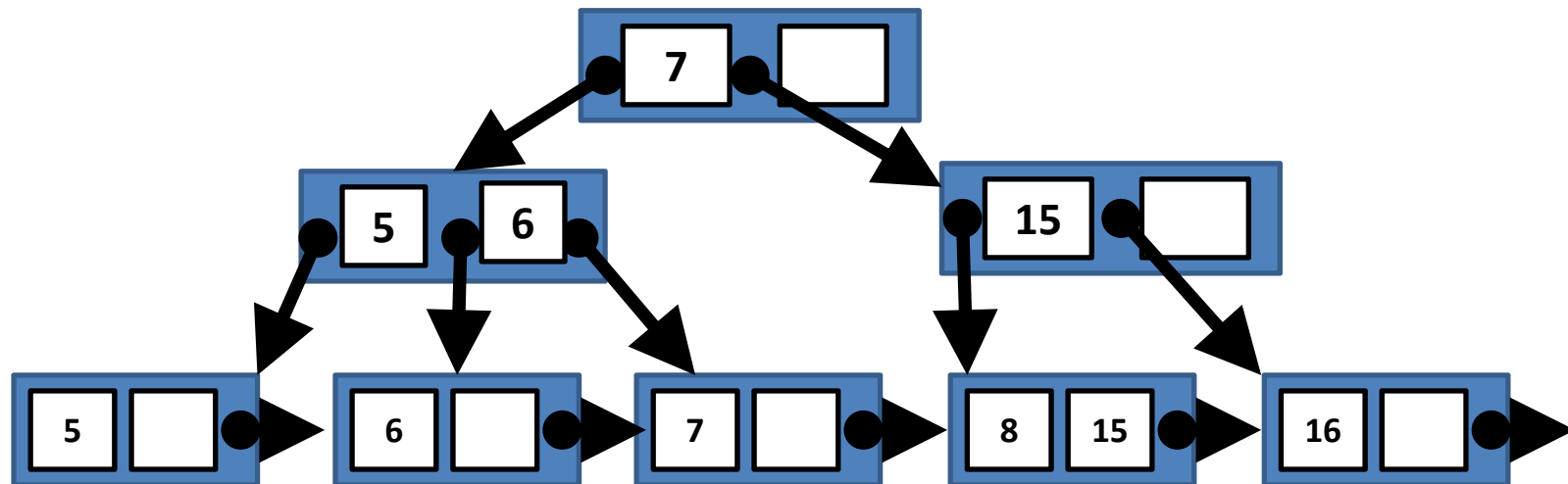


Obrisati 4





Kada se mešaju 5 i 6, gubi se ključ u roditelju koji ih razdvaja. Zato nestaje 5 iz korena.



Zadatak:

Baza podataka sadrži 3 000 000 slogova dužine 250 B. Smatrati da slogovi nisu uređeni ni po jednom kriterijumu. Veličina bloka na disku je 4096 B, veličina ključa 15 B, a veličina pokazivača na blok je 6 B. Odrediti:

- koliko MB na disku zauzima osnovna datoteka, (3 poena),
- ako se za indeksiranje koristi B stablo, odrediti koliko puta se ubrzava pristup i za koliko procenata se poveća datoteka (7 poena).

Rešenje:

Veličina datoteke je:

$3\,000\,000 * 250B = 750\,000\,000\,B = 750\,MB$ (kako računaju proizvođači diskova)

Ili 715.256 MB, ako smatramo da je $1kB = 1024B$

... i to donosi ...

0 poena !!!

Baza podataka sadrži 3 000 000 slogova dužine 250 B. Smatrati da slogovi nisu uređeni ni po jednom kriterijumu. Veličina bloka na disku je 4096 B, veličina ključa 15 B, a veličina pokazivača na blok je 6 B. Odrediti:

- **koliko MB na disku zauzima osnovna datoteka, (3 poena),**
- ako se za indeksiranje koristi B stablo, odrediti koliko puta se ubrzava pristup i za koliko procenata se poveća datoteka (7 poena).

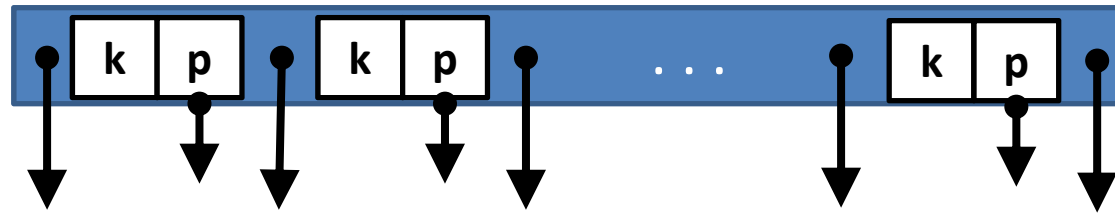
$$fb = \left\lceil \frac{4096}{250} \right\rceil = \lceil 16.384 \rceil = 17$$

$$bb = \left\lceil \frac{r}{fb} \right\rceil = \left\lceil \frac{3.000.000}{17} \right\rceil = 176471$$

Veličina_datoteke je = $bb * \text{veličina_bloka} = 176471 * 4096 \text{ B} = 724 925 408 \text{ B} = 724.92 \text{ MB}$

Baza podataka sadrži 3 000 000 slogova dužine 250 B. Smatrati da slogovi nisu uređeni ni po jednom kriterijumu. Veličina bloka na disku je 4096 B, veličina ključa 15 B, a veličina pokazivača na blok je 6 B. Odrediti:

- koliko MB na disku zauzima osnovna datoteka, (3 poena),
- **ako se za indeksiranje koristi B stablo, odrediti koliko puta se ubrzava pristup i za koliko procenata se poveća datoteka (7 poena).**



Ako je **k** ključ, a **p** pokazivač, i ako sa **q** označimo broj parova (k,p), onda je veličina čvora jednaka

$$V \geq q(k+p) + (q+1)p$$

$$V \geq qk + qp + qp + p$$

$$V \geq q(2p+k) + p$$

$$4096 \geq q(2 \cdot 6 + 15) + 6$$

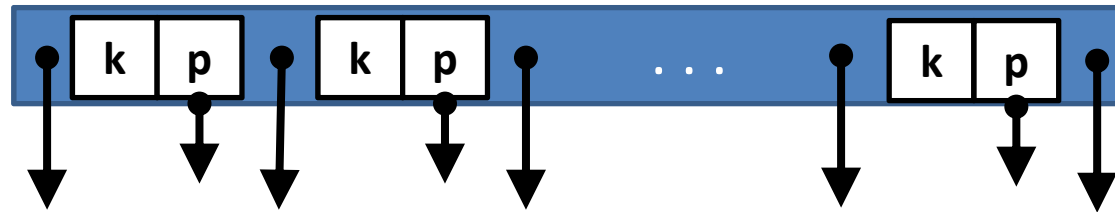
$$q \leq 4090 / 27 = 151.48$$

$$\mathbf{q = 151}$$

Dakle, jednim čvorom možemo indeksirati 151 slog datoteke. Koliko čvorova nam je potrebno da bismo indeksirali celu datoteku?

Pošto datoteka nije uređena, potreban nam je, praktično, sekundarni indeks, pa je

$$\text{broj_listova} = B_0 = \lceil 3.000.000 / 151 \rceil = \lceil 19867.55 \rceil = 19868$$



Na koliko listova ukazuje čvor iz narednog nivoa?

$$q+1$$

Dakle, za 19868 listova, potrebno je:

$$B_1 = \lceil 19868 / 152 \rceil = \lceil 130.71 \rceil = 131$$

Proces se nastavlja sve dok ne dođemo do korena. Ovde je to već na narednom nivou.

$$B_2 = \lceil 131 / 152 \rceil = \lceil 0.86 \rceil = 1$$

Koliko puta se ubrzava pristup?

Da nemamo B-stablo, pristupali bi sekvencijalno. Imamo 187500 blokova. U proseku nam treba $187500/2 = 93750$ pristupa.

Ako koristimo B-stablo, koje ima 3 nivoa, broj pristupa je 4.

Pa je ubrzanje $93750 / 4 = 23437.5$ puta!

A povećanje datoteke je za veličinu B-stabla. Ono ima $19868 + 131 + 1 = 20000$ blokova.

Kako osnovna datoteka ima 187500 blokova, dodavanjem indeksa, tj. B-stabla, datoteku smo povećali $2000 / 187500 = 0.1067 \approx 10.7\%$

Odgovor na zadatak je, dakle, **ubrzanje je 23437.5 puta, a povećanje datoteke je 10.7%.**

Neki će reći da ovo nije tačno rešenje, zato što smo zanemarili unutrašnje čvorove koji takođe sadrže pokazivače na blokove sa podacima.

U prethodnom slučaju, broj unutrašnjih čvorova je 132. Oni mogu indeksirati $132 * 151 = 19781$ slogova primarne datoteke. Izgleda kao strahovita greška.

Ali...

Imali smo **19868** listova. Unutrašnji čvorovi čine tek **0.66%** broja listova.

Broj listova bi trebalo smanjiti za manje od 0.66%, jer se smanjenjem broja listova, smanjuje i broj unutrašnjih čvorova.

Potrebno je mnogo više izračunavanja da bi se dobila tačna vrednost.

Da li će to promeniti broj nivoa stabla?

VRLO NEVEROVATNO!!!

Da li će se promeniti veličina datoteke?

Hoće, ali tek na nekoj decimali !!!

Da li ima smisla računati tačan broj čvorova u B-stablu?

Zaključak donesite sami !!!

Umesto $1+131+19868$ šeme, osnovnu datoteku bi mogli da indeksiramo sa $1+130+19737$ čvorova (potrebno je da ukupno imamo 19868 čvorova). Ukupan broj blokova u B-stablu je 19868, pa je uvećanje datoteke 10.596267%. Ranijim postupkom smo dobili 10.666667% i napravili grešku od 0.66%.

Zadatak:

Birački spisak jednog grada sadrži podatke o 2 000 000 građana sa pravom glasa. Ako se o svakom građaninu pamte sledeći podaci: JMBG (14B), br. lične karte (7B), ime (30B), prezime(30B), godina rođenja (4B), adresa (50B). Smatrati da slogovi nisu uređeni ni po jednom kriterijumu. Veličina bloka na disku je 2048B, a veličina pokazivača na blok je 6B. Odrediti:

- koliko MB na disku zauzima osnovna datoteka, (2 poena),
- ako se za indeksiranje koristi B+ stablo, čiji su čvorovi popunjeni 75%, odrediti koliko puta se ubrzava pristup i za koliko procenata se poveća datoteka (6 poena).

Rešenje:

Veličina sloga je: $14+7+30+30+4+50 = 135$

$$fb = \left\lfloor \frac{2048}{135} \right\rfloor = \lfloor 15.17 \rfloor = 15$$

$$bb = \left\lceil \frac{r}{fb} \right\rceil = \left\lceil \frac{2.000.000}{15} \right\rceil = \lceil 133333.33 \rceil = 133334$$

Veličina datoteke je: $133334 * 2048 \text{ B} = 273068032 \text{ B} = 260.41 \text{ MB}$

Zadatak:

...

- ako se za indeksiranje koristi B+ stablo, čiji su čvorovi popunjeni 75%, odrediti koliko puta se ubrzava pristup i za koliko procenata se poveća datoteka (6 poena).

Kako izgleda list B+ stabla?



Ako je **k** ključ, a **p** pokazivač, i ako sa **q** označimo broj parova (k,p), onda je veličina čvora jednaka

$$\begin{aligned} V &\geq q(k+p) + p \\ 2048 &\geq q(14+6) + 6 \\ q &\leq 2042 / 20 = 102.1 \\ \mathbf{q} &= \mathbf{102} \end{aligned}$$

Dakle, da je list 100% popunjen, on bi imao 102 pokazivača na slogove izvorne datoteke. Ali, on je „samo“ 75% popunjen. Dakle, umesto 102, svaki list će ukazivati na:

$$q' = 102 * 0.75 = 76.6$$

Dogovorićemo se da zaokružimo vrednost, pa je:

$$q' = 77$$

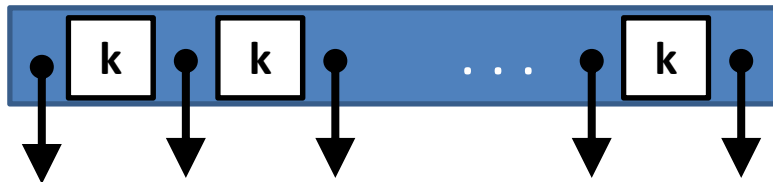
Sada treba videti koliko listova je potrebno za indeksiranje svih slogova osnovne datoteke.

$$\text{broj_listova} = B_0 = \lceil 2.000.000 / 77 \rceil = \lceil 25974.025 \rceil = 25975$$

Koliko je potrebno čvorova na drugom nivou?

To ne možemo odmah izračunati, jer se struktura unutrašnjih čvorova razlikuje od listova.

Kako izgleda unutrašnji čvor B+ stabla?



Ako je **k** ključ, a **p** pokazivač, i ako sa **q** označimo broj pokazivača na podstabla

$$V \geq (q-1)k + qp$$

$$2048 \geq qk - k + qp$$

$$2048 \geq q(k+p) - k$$

$$q \leq 2062 / 20 = 103.1$$

$$\mathbf{q = 103}$$

Kako je popunjenost 75%:

$$\mathbf{q'' = 103 * 0.75 = 77}$$

Ako je broj listova 25975, na sledećem nivou potrebno je:

$$B_1 = \lceil 25975 / 77 \rceil = \lceil 337.34 \rceil = 338$$

Proces se nastavlja sve dok ne dođemo do korena.

$$B_2 = \lceil 338 / 77 \rceil = \lceil 4.4 \rceil = 5$$

$$B_3 = \lceil 5 / 77 \rceil = \lceil 0.065 \rceil = 1$$

Ubrzanje pristupa je: $133334 / (2 * 5) = \mathbf{13333.4}$ puta

A povećanje datoteke: $(25975 + 338 + 5 + 1) / 133334 \approx \mathbf{19.7\%}$