

Indeksne datoteke s B-stablima

B-stablo, formiranje, traženje, ažuriranje i obrada
indeksne datoteke sa B-stabлом

Sadržaj

- ▶ Osnovno B-stablo
- ▶ Formiranje datoteke s B-stablom
- ▶ Traženje u datoteci s B-stablom
- ▶ Obrada datoteke s B-stablom
- ▶ Ažuriranje datoteke s B-stablom
- ▶ Ocena karakteristika datoteka s B-stablom
- ▶ B^* -stablo, $B^{\#}$ -stablo, B^+ -stablo
- ▶ Indeks-sekvencijalna datoteka s B-stablom

Osnovno B-stablo

► **B-stablo**

- ▶ puno stablo i stablo traženja
- ▶ gusto popunjeni, dinamički indeks
- ▶ visine h
 - ▶ svi listovi su na jednakoj udaljenosti od korena
 - ▶ put od korena do bilo kog lista je iste dužine
- ▶ ranga r ($r \geq 2$)
 - ▶ uvodi ograničenje na dozvoljeni broj elemenata u svakom čvoru
- ▶ reda $n = 2r + 1$

Osnovno B-stablo

► **B-stablo**

- ▶ čvor = blok zone indeksa
- ▶ čvor sadrži niz elemenata
- ▶ element = propagirana vrednost ključa iz primarne zone
- ▶ svaki čvor sadrži maksimalno $2r$ elemenata
- ▶ svaki čvor, izuzev korena, sadrži minimalno r elemenata
- ▶ koren sadrži minimalno 1 element
- ▶ svaki čvor sa m elemenata, koji ne predstavlja list, poseduje $m + 1$ direktno podređenih čvorova

Osnovno B-stablo

► Format čvora B-stabla

- niz elemenata
- svaki element: trojka (k_e, A_e, P_e) , $e \in \{1, \dots, m\}$
 - k_e vrednost ključa sloga S_i ($i \in \{1, 2, \dots, N\}$)
 - A_e pridruženi podatak
 - P_e pokazivač ka podstablu s većim vrednostima ključa od k_e

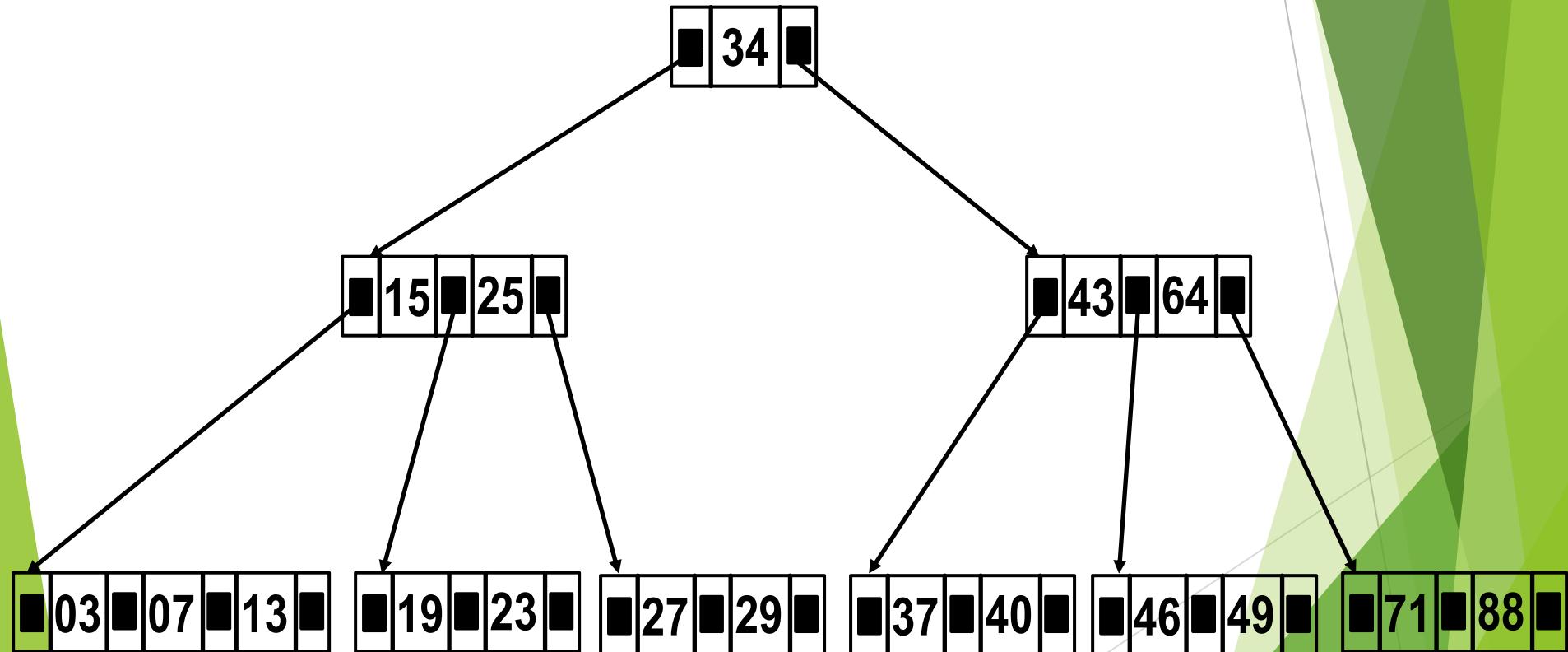
Zaglavlje bloka	P_0	k_1	A_1	P_1	k_2	A_2	P_2	...	k_m	A_m	P_m	Neiskorišćeni prostor

► Uslovi stabla traženja

- $(\forall i \in \{1, \dots, m - 1\})(k_i < k_{i+1})$
- $(\forall k \in K(P_0))(k < k_1)$
- $(\forall i \in \{1, \dots, m - 1\})(\forall k \in K(P_i))(k_i < k < k_{i+1})$
- $(\forall k \in K(P_m))(k_m < k)$

Osnovno B-stablo

- ▶ Primer:
 - ▶ B-stablo sa $N = 18$ slogova, ranga $r = 2$, visine $h = 3$



Osnovno B-stablo

- ▶ Popunjenošt B-stabla
 - ▶ za isti broj slogova N i rang r , B-stablo može posedovati različite visine i različite brojeve čvorova
- ▶ Ekstremni slučajevi popunjenošti B-stabla
 - ▶ **poluprazno (polupuno) B-stablo**
 - ▶ svi čvorovi, osim korena, sadrže po r elemenata
 - ▶ koren sadrži samo 1 element
 - ▶ stablo ne može biti manje popunjeno od polupraznog
 - ▶ **kompletno (popunjeno) B-stablo**
 - ▶ svi čvorovi sadrže po $2r$ elemenata
 - ▶ stablo ne može biti više popunjeno od kompletognog

Osnovno B-stablo

- ▶ Broj čvorova i elemenata
 - ▶ na različitim nivoima hijerarhije B-stabla ranga r

Nivo	Visina	Kompletno B-stablo		Poluprazno B-stablo	
		Broj čvorova	Broj elemenata	Broj čvorova	Broj elemenata
0	1	1	$2r$	1	1
1	2	$(2r + 1)^1$	$2r(2r + 1)^1$	2	$2r$
2	3	$(2r + 1)^2$	$2r(2r + 1)^2$	$2(r + 1)^1$	$2r(r + 1)^1$
...
$i - 1$	i	$(2r + 1)^{i-1}$	$2r(2r + 1)^{i-1}$	$2(r + 1)^{i-2}$	$2r(r + 1)^{i-2}$
...
$h - 1$	h	$(2r + 1)^{h-1}$	$2r(2r + 1)^{h-1}$	$2(r + 1)^{h-2}$ ^{8 / 75}	$2r(r + 1)^{h-2}$

Osnovno B-stablo

- ▶ Broj čvorova stabla
 - ▶ kompletno stablo

$$C_{kp} = \sum_{i=1}^h (2r+1)^{i-1} = \sum_{i=0}^{h-1} (2r+1)^i = \frac{(2r+1)^h - 1}{2r}$$

- ▶ poluprazno stablo

$$C_{pp} = 1 + 2 \sum_{i=1}^{h-1} (r+1)^{i-1} = 1 + 2 \sum_{i=0}^{h-2} (r+1)^i = 1 + 2 \frac{(r+1)^{h-1} - 1}{r}$$

Osnovno B-stablo

► Broj čvorova i visina stabla

- kompletno stablo

$$C_{kp} = \frac{(2r+1)^h - 1}{2r}, \quad N = 2rC_{kp} \Rightarrow N = (2r+1)^h - 1$$

$$h_{kp} = \log_{2r+1}(N+1), \quad h_{min} = \lceil \log_{2r+1}(N+1) \rceil$$

$$C_{kp} = \frac{N}{2r}, \quad C_{min} = \left\lceil \frac{N}{2r} \right\rceil$$

- visina ne može biti manja od h_{min}
- ukupan broj čvorova ne može biti manji od C_{min}

Osnovno B-stablo

► Broj čvorova i visina stabla

- ▶ poluprazno stablo

$$C_{pp} = 1 + 2 \frac{(r+1)^{h-1} - 1}{r}, \quad N = 1 + r(C_{pp} - 1) \Rightarrow N = 2(r+1)^{h-1} - 1$$

$$h_{pp} = 1 + \log_{r+1} \frac{N+1}{2}, \quad h_{max} = 1 + \left\lfloor \log_{r+1} \frac{N+1}{2} \right\rfloor$$

$$C_{pp} = 1 + \frac{N-1}{r}, \quad C_{max} = 1 + \left\lfloor \frac{N-1}{r} \right\rfloor$$

- ▶ visina ne može biti veća od h_{max}
- ▶ ukupan broj čvorova ne može biti veći od C_{max}

Osnovno B-stablo

► Broj čvorova i visina stabla

► $h_{min} \leq h \leq h_{max}, C_{min} \leq C \leq C_{max}$

N	r	h_{min}	h_{max}	C_{min}	C_{max}
10^3	50	2	2	10	20
10^4	50	2	3	10^2	$2 \cdot 10^2$
10^5	50	3	3	10^3	$2 \cdot 10^3$
10^6	50	3	4	10^4	$2 \cdot 10^4$

$$h_{min} = \lceil \log_{2r+1}(N+1) \rceil \quad h_{max} = 1 + \left\lfloor \log_{r+1} \frac{N+1}{2} \right\rfloor$$

$$C_{min} = \left\lceil \frac{N}{2r} \right\rceil$$

$$C_{max} = 1 + \left\lfloor \frac{N-1}{r} \right\rfloor$$

Sadržaj

- ▶ Osnovno B-stablo
- ▶ Formiranje datoteke s B-stabлом
- ▶ Traženje u datoteci s B-stabлом
- ▶ Obrada datoteke s B-stabлом
- ▶ Ažuriranje datoteke s B-stabлом
- ▶ Ocena karakteristika datoteka s B-stabлом
- ▶ B*-stablo, B[#]-stablo, B⁺-stablo
- ▶ Indeks-sekvencijalna datoteka s B-stabлом

Formiranje datoteke s B-stablim

► Indeksna metoda pristupa

- ▶ operativni sistem
 - ▶ OS mainframe računara poseduju metode pristupa za formiranje, korišćenje i ažuriranje indeksnih datoteka sa *B*-stablima
- ▶ programski jezici
 - ▶ korišćenje specijalizovanih biblioteka za indeksnu metodu pristupa, ili
 - ▶ korisnici sami pišu svoje metode
- ▶ SUBP
 - ▶ poseduju sopstvene indeksne metode pristupa
 - ▶ koriste ih u izgradnji fizičkih struktura baza podataka

Formiranje datoteke s B-stablim

► **Struktura datoteke s B-stablim**

- ▶ **zona indeksa: spregnuta struktura, B-stablo**
 - ▶ dinamički, gusto popunjeni indeks
 - ▶ svaka vrednost ključa primarne zone propagira se u zonu indeksa
 - ▶ dinamičko ažuriranje - prati ažuriranje primarne zone
 - ▶ omogućava traženja u primarnoj zoni
- ▶ **primarna zona: serijska datoteka**
 - ▶ iskorišćenje dobrih osobina serijske datoteke pri ažuriranju
 - ▶ pod pretpostavkom da se traženja ne vrše direktno u serijskoj datoteci

Formiranje datoteke s B-stablon

► **Formiranje datoteke s B-stablon**

► **formiranje primarne zone**

- ▶ preuzimanjem postojeće (serijske) datoteke ili
- ▶ sukcesivnim učitavanjem slogova ulazne (serijske) datoteke

► **formiranje zone indeksa**

- ▶ upisivanje zaglavlja stabla traženja u zonu indeksa
 - ▶ inicijalno, nedefinisane vrednosti pokazivača na koren stabla i krajnji levi čvor stabla
 - ▶ inicijalizuje se pokazivač na prvi blok u lancu praznih blokova
- ▶ formiranje lanca praznih blokova (čvorova)
- ▶ sukcesivno čitanje slogova primarne zone i formiranje B-stabla, dinamičkim upisivanjem novih elemenata
- ▶ upisivanje prvog elementa ⇒ formiranje korena stabla

Formiranje datoteke s B-stablon

► **Upis novog elementa u B-stablo**

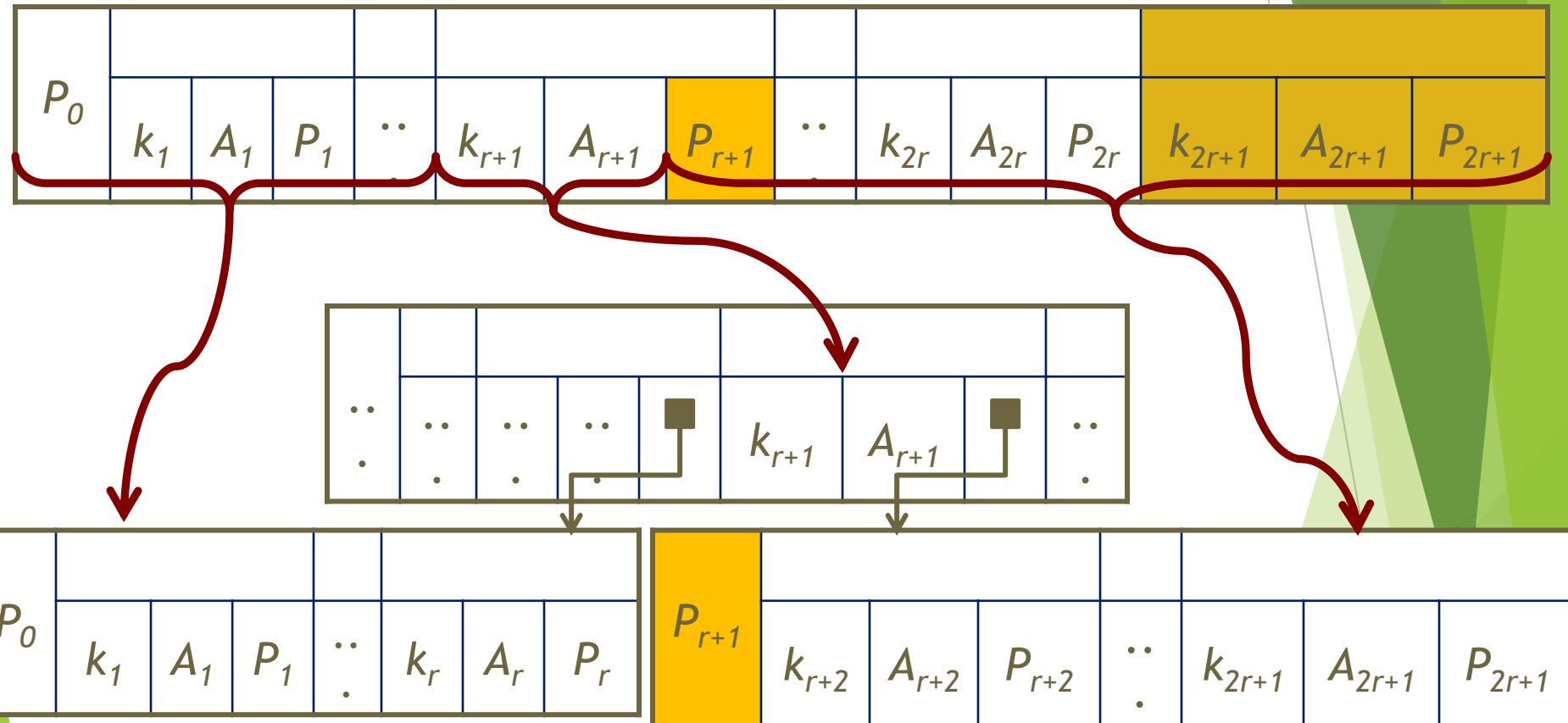
- ▶ prethodi mu neuspešno traženje elementa (tso)
 - ▶ započinje u korenu
 - ▶ uvek završava u listu
 - ▶ upoređivanjem argumenta traženja s vrednostima ključa u svakom čvoru i
 - ▶ praćenjem pokazivača na čvorove na jednom pristupnom putu od korena do lista
- ▶ započinje na mestu zaustavljanja neuspešnog traženja u listu
- ▶ list može biti potpuno ili delimično popunjeno

Formiranje datoteke s B-stablon

► Upis novog elementa u B-stablo

- ▶ aktuelni broj elemenata u listu m_e - dve mogućnosti
- ▶ (A) delimično popunjen list: $m_e < 2r$
 - ▶ novi element upisuje se na poziciju zaustavljanja traženja
 - ▶ elementi s većom vrednošću ključa u čvoru pomeraju se za jednu poziciju udesno
- ▶ (B) potpuno popunjen list: $m_e = 2r$
 - ▶ podela lista na dva lista - alocira se novi, desni list
 - ▶ formira se u OM uređeni niz od $2r + 1$ elemenata
 - ▶ prva polovina niza ostaje u levom (postojećem) listu:
 - ▶ srednji element upisuje se u nadređeni čvor - može izazvati deljenje nadređenog čvora
 - ▶ druga polovina niza upisuje se u desni (novi) list
 - ▶ ukoliko se deli koren, dolazi do formiranja novog korena i povećavanja visine stabla za 1

Formiranje datoteke s B-stabljom



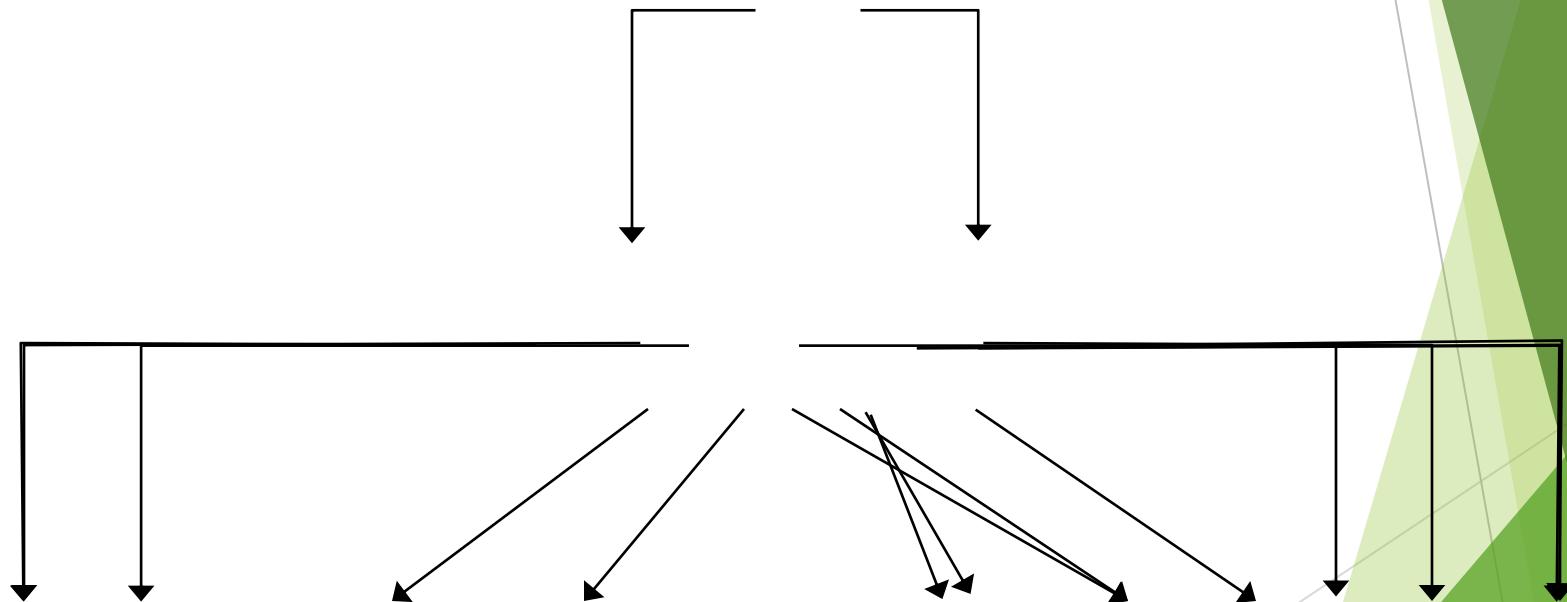
Formiranje datoteke s B-stabлом

34	07	03	15	19	29	64	43	23	27	13	49	25	71	88	37	46	40
----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----

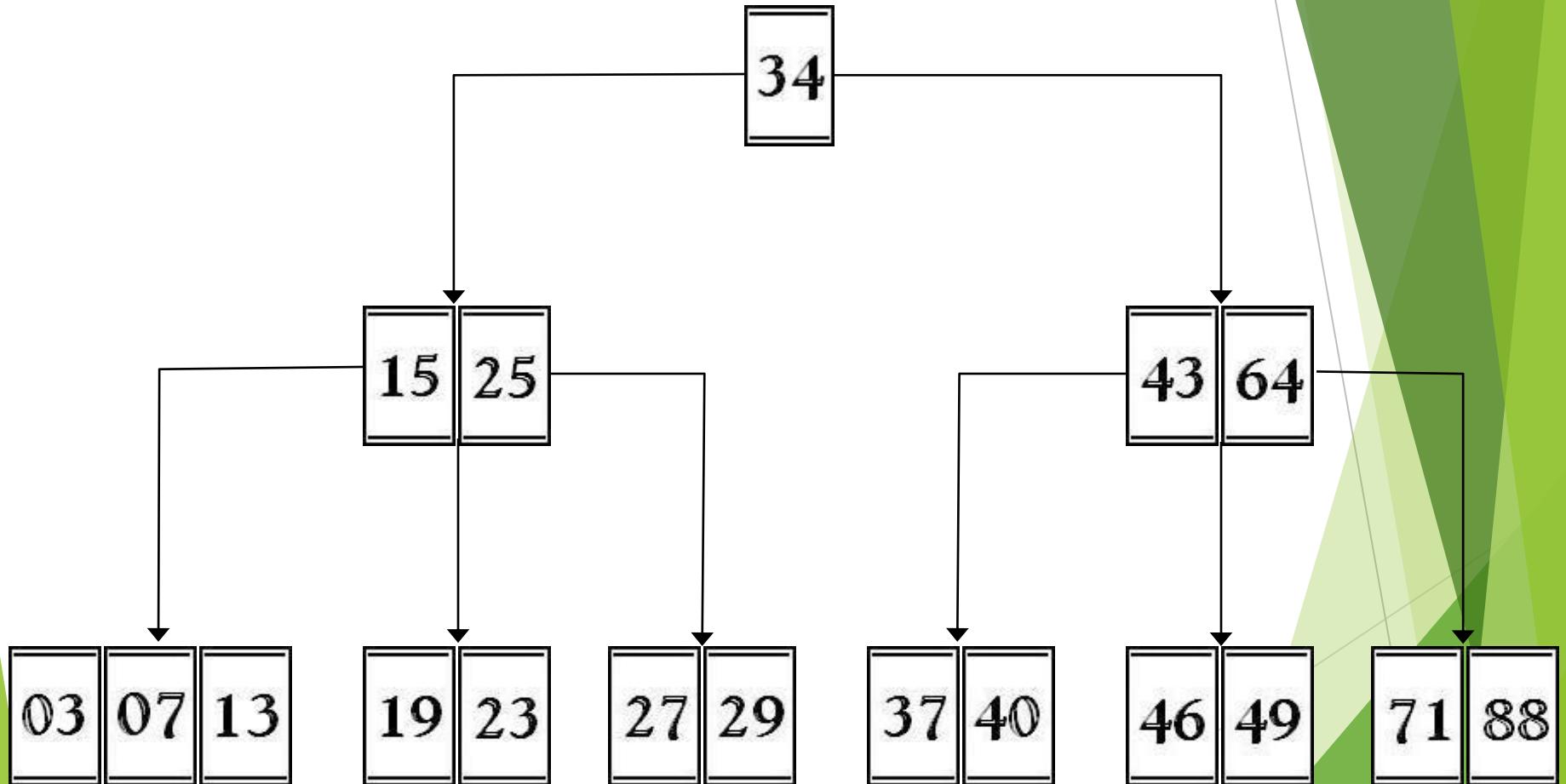
- ▶ $N = 18$ slogova
- ▶ rang stabla: $r = 2$
- ▶ visina stabla: $h = 3$
- ▶ kapacitet svakog čvora: $2r = 4$ elementa

Formiranje datoteke s B-stablotom

34	07	03	15	19	29	64	43	23	27	13	49	25	71	88	37	46	40
----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----



Formiranje datoteke s B-stablotom



Sadržaj

- ▶ Osnovno B-stablo
- ▶ Formiranje datoteke s B-stablon
- ▶ Traženje u datoteci s B-stablon
- ▶ Obrada datoteke s B-stablon
- ▶ Ažuriranje datoteke s B-stablon
- ▶ Ocena karakteristika datoteka s B-stablon
- ▶ B*-stablo, B[#]-stablo, B⁺-stablo
- ▶ Indeks-sekvencijalna datoteka s B-stablon

Traženje u datoteci s B-stablim

► Traženje logički narednog sloga

- ▶ modifikovani simetrični postupak prolaska kroz B-stablo
 - ▶ naizmenično pristupanje listovima i njihovim nadređenim elementima
- ▶ vrši se od tekućeg elementa stabla
 - ▶ inicijalno, tekući element stabla je element sa najmanjom vrednošću ključa u krajnjem levom listu
- ▶ upoređuju se argument traženja a i vrednosti ključa elemenata stabla k_e
 - ▶ traženje se uspešno završava kada je $a = k_e$
 - ▶ neuspešno traženje završava se nailaskom na element sa $k_e > a$ ili nailaskom na kraj krajnjeg desnog lista

Traženje u datoteci s B-stablu

► Traženje slučajno odabranog sloga

- traženje započinje u korenu stabla
 - eventualno se nastavlja u podređenim čvorovima
 - završava se u čvoru u kojem je element pronađen ili u listu
 - na svakom nivou hijerarhije stabla pristupa se najviše jednom čvoru
- upoređuju se argument traženja a i vrednosti ključa elemenata stabla k_e
- ako se pri traženju nađe element, za koji je $a = k_e$
 - traženom slogu pristupa se na osnovu adrese A_e
- ako se, pri traženju u čvoru, dođe do elementa sa $k_e > a$
 - traženje se nastavlja u odgovarajućem podstablu

Traženje u datoteci s B-stablolom

► Traženje slučajno odabranog sloga

- uspešno, završava u bilo kojem čvoru
 - broj pristupa, samo jedan bafer u OM za stablo pristupa

$$R_u = h + 1$$

- broj pristupa, h bafera u OM za stablo pristupa

$$0 \leq R_u \leq h + 1$$

- neuspešno, završava uvek u listu

- broj pristupa, samo jedan bafer u OM za stablo pristupa

$$R_n = h$$

- broj pristupa, h bafera u OM za stablo pristupa

$$0 \leq R_n \leq h$$

- celo stablo u OM: $R_u = R_n = 0$

Sadržaj

- ▶ Osnovno B-stablo
- ▶ Formiranje datoteke s B-stablolom
- ▶ Traženje u datoteci s B-stablolom
- ▶ Obrada datoteke s B-stablolom
- ▶ Ažuriranje datoteke s B-stablolom
- ▶ Ocena karakteristika datoteka s B-stablolom
- ▶ B*-stablo, B[#]-stablo, B⁺-stablo
- ▶ Indeks-sekvencijalna datoteka s B-stablolom

Obrada datoteke s B-stabлом

- ▶ Može se koristiti
 - ▶ i kao vodeća i kao obrađivana
 - ▶ i u režimu redosledne i u režimu direktne obrade
 - ▶ u svim slučajevima, pokazuje solidne performanse

Obrada datoteke s B-stablon

- ▶ Ukupan broj pristupa datoteci R_{uk}^v - parametri
 - ▶ da li je stablo kompletno ili poluprazno
 - ▶ broj čvorova kompletogn stabla je najmanji mogući, a polupraznog najveći mogući
 - ▶ da li je rezervisano h bafera za ceo pristupni put ili samo 1 bafer za čvorove B-stabla
 - ▶ u slučaju h bafera, ceo pristupni put staje u OM pa se svakom čvoru stabla pristupa tačno jedanput
 - ▶ u slučaju jednog bafera, svakom neterminalnom čvoru mora se pristupati barem onoliko puta koliko elemenata poseduje
 - ▶ da li su sukcesivno traženi slogovi u primarnoj zoni uvek smešteni u fizički susedne lokacije, ili uvek u lokacije različitih blokova

Obrada datoteke s B-stablon

- ▶ Ukupan broj pristupa datoteci R_{uk}^v
 - ▶ uloga vodeće datoteke

$$C + \left\lceil \frac{N}{f} \right\rceil \leq R_{uk}^v \leq C_h + E_{1 \rightarrow h-1} + N$$

$$C_{min} + \left\lceil \frac{N}{f} \right\rceil \leq R_{uk}^v \leq C_h^{max} + E_{1 \rightarrow h-1}^{max} + N$$

$$\left\lceil \frac{N}{2r} \right\rceil + \left\lceil \frac{N}{f} \right\rceil \leq R_{uk}^v \leq \left\lfloor \frac{N+1}{r+1} \right\rfloor + \left\lfloor \frac{N-r}{r+1} \right\rfloor + N$$

Obrada datoteke s B-stablom

► Redosledna obrada

- putem vodeće datoteke sa $N_v = N_v^u + N_v^n$ slogova
 - N_v^u slogova koji iniciraju uspešno traženje
 - N_v^n slogova koji iniciraju neuspešno traženje
- ukupan broj pristupa datoteci kao obrađivanoj, R_{uk}^r

$$\left\lceil \frac{N}{2r} \right\rceil + \left\lceil \frac{N_v^u}{f} \right\rceil \leq R_{uk}^r \leq \left\lfloor \frac{N+1}{r+1} \right\rfloor + \left\lfloor \frac{N-r}{r+1} \right\rfloor + N_v^u$$

- očekivani broj pristupa pri uspešnom ili neuspešnom traženju jednog logički narednog sloga

$$\bar{R} = \frac{R_{uk}^r}{N_v}$$

Obrada datoteke s B-stablon

► Direktna obrada

- putem vodeće datoteke sa $N_v = N_v^u + N_v^n$ slogova
 - N_v^u slogova koji iniciraju uspešno traženje
 - N_v^n slogova koji iniciraju neuspešno traženje
- relativno brz pristup slučajno odabranom slogu
- ukupan broj pristupa R_{uk}^d

$$R_{uk}^d = R_u N_v^u + R_n N_v^n$$

Sadržaj

- ▶ Osnovno B-stablo
- ▶ Formiranje datoteke s B-stablotom
- ▶ Traženje u datoteci s B-stablotom
- ▶ Obrada datoteke s B-stablotom
- ▶ Ažuriranje datoteke s B-stablotom
- ▶ Ocena karakteristika datoteka s B-stablotom
- ▶ B*-stablo, B[#]-stablo, B⁺-stablo
- ▶ Indeks-sekvencijalna datoteka s B-stablotom

Ažuriranje datoteke s B-stablon

- ▶ **Upis novog i brisanje postojećeg sloga**
 - ▶ vrši se u režimu direktne obrade
 - ▶ izvodi ga metoda pristupa na zahtev aplikativnog programa
 - ▶ vrši se upis, odnosno logičko brisanje elementa iz primarne zone i odgovarajućeg elementa iz stabla pristupa
- ▶ **Upis novog elementa u B-stablo**
 - ▶ na isti način, kao u slučaju formiranja B-stabla

Ažuriranje datoteke s B-stablon

- ▶ **Brisanje elementa iz B-stabla**
 - ▶ prethodi mu uspešno traženje
 - ▶ logičko brisanje sloga u primarnoj zoni
 - ▶ fizičko brisanje elementa iz B-stabla
 - ▶ element se sme fizički izbrisati jedino ako se nalazi u listu

Ažuriranje datoteke s B-stablon

► Brisanje elementa iz B-stabla

- (A) element za brisanje nalazi se u listu
 - (A1) list sadrži $m_e > r$ elemenata ili je koren stabla
 - fizičko oslobođanje lokacije izbrisaniog elementa
 - po potrebi, pomeranje ostalih elemenata u listu za jednu poziciju uлево
 - nakon brisanja, u listu ostaje $m_e - 1 \geq r$ elemenata
 - (A2) list sadrži $m_e = r$ elemenata i nije koren stabla
 - fizičko oslobođanje lokacije izbrisaniog elementa nije dozvoljeno
 - svaki čvor, osim korena, mora imati minimalno r elemenata
 - (A21) postoji barem jedan susedni čvor sa $m_f > r$ elemenata
 - primena **tehnike pozajmljivanja elemenata** iz susednog čvora
 - (A22) svi susedni čvorovi imaju $m_f = r$ elemenata
 - primena **tehnike spajanja dva čvora** u jedan

Ažuriranje datoteke s B-stablon

► Brisanje elementa iz B-stabla

► susedni čvorovi

- čvorovi C_e i C_f su susedni, ako imaju zajednički direktno nadređeni čvor C i ako važi

$$(\exists ! k \in K(C))(k_m < k < k_1)$$

- $K(C)$ - skup vrednosti ključa u čvoru C
- k_m - najveća vrednost ključa u čvoru C_e
- k_1 - najmanja vrednost ključa u čvoru C_f

Ažuriranje datoteke s B-stablon

► Brisanje elementa iz B-stabla

► tehniku pozajmljivanja elemenata od suseda

- ▶ formira se u OM uređeni niz od $m_f + r - 1 + 1$ elemenata
 - ▶ svi elementi susednog čvora, jedan nadređeni i $r - 1$ element čvora iz kojeg se briše element
 - ▶ $\lfloor (m_f + r) / 2 \rfloor$ elemenata smešta se u levi susedni čvor
 - ▶ $\lfloor (m_f + r) / 2 \rfloor + 1$ -vi element smešta se u nadređeni čvor
 - ▶ preostali elementi smeštaju se u desni susedni čvor

► tehniku spajanja dva susedna čvora u jedan

- ▶ formira se u OM uređeni niz od $r + r - 1 + 1$ elemenata
 - ▶ r elemenata susednog čvora, jedan nadređeni i $r - 1$ element čvora iz kojeg se briše element
 - ▶ svih $2r$ elemenata smešta se u levi susedni čvor
 - ▶ desni susedni čvor postaje prazan - dealocira se
 - ▶ fizičko brisanje elementa iz nadređenog čvora

Ažuriranje datoteke s B-stablon

- ▶ **Brisanje elementa iz B-stabla**
 - ▶ **tehnika spajanja dva susedna čvora u jedan**
 - ▶ fizičko brisanje elementa iz nadređenog čvora
 - ▶ može izazvati novo pozajmljivanje ili spajanje čvorova, na višem nivou u stablu
 - ▶ u ekstremnom slučaju, može doći do spajanja jedina dva direktno podređena čvora korenu
 - ▶ izaziva dealociranje korena i smanjenje visine stabla za jedan

Ažuriranje datoteke s B-stablon

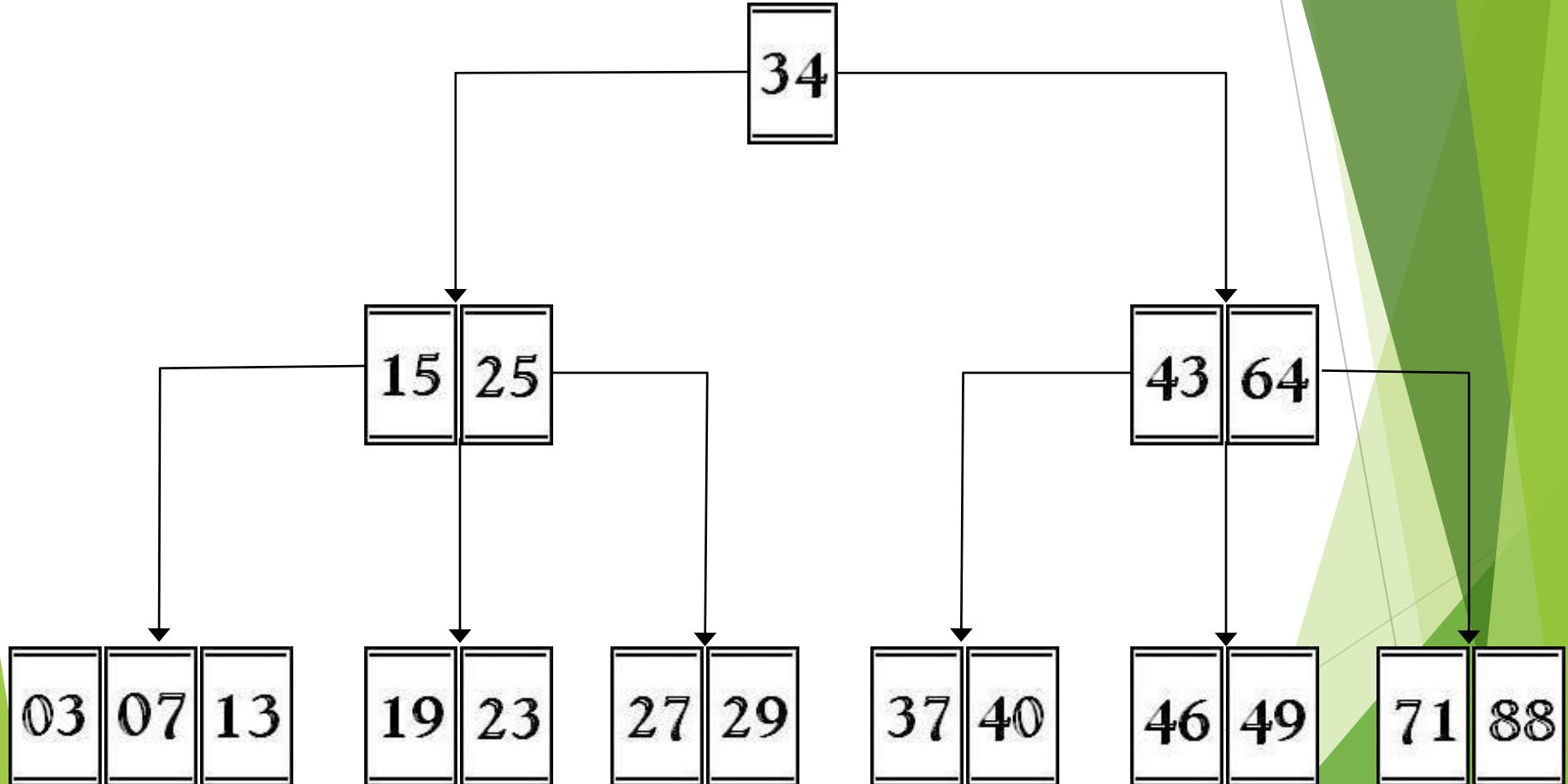
► Brisanje elementa iz B-stabla

► (B) element za brisanje ne nalazi se u listu

- zamena elementom koji sadrži prvu veću vrednost ključa
- predstavlja krajnji levi element u krajnjem levom čvoru desnog podstabla u odnosu na element koji se briše
- nakon zamene (upisivanja u lokaciju izbrisanih element), element s prvom većom vrednošću ključa fizički se briše iz lista
- povratak na slučaj (A) - primena za fizičko brisanje elementa s prvom većom vrednošću ključa

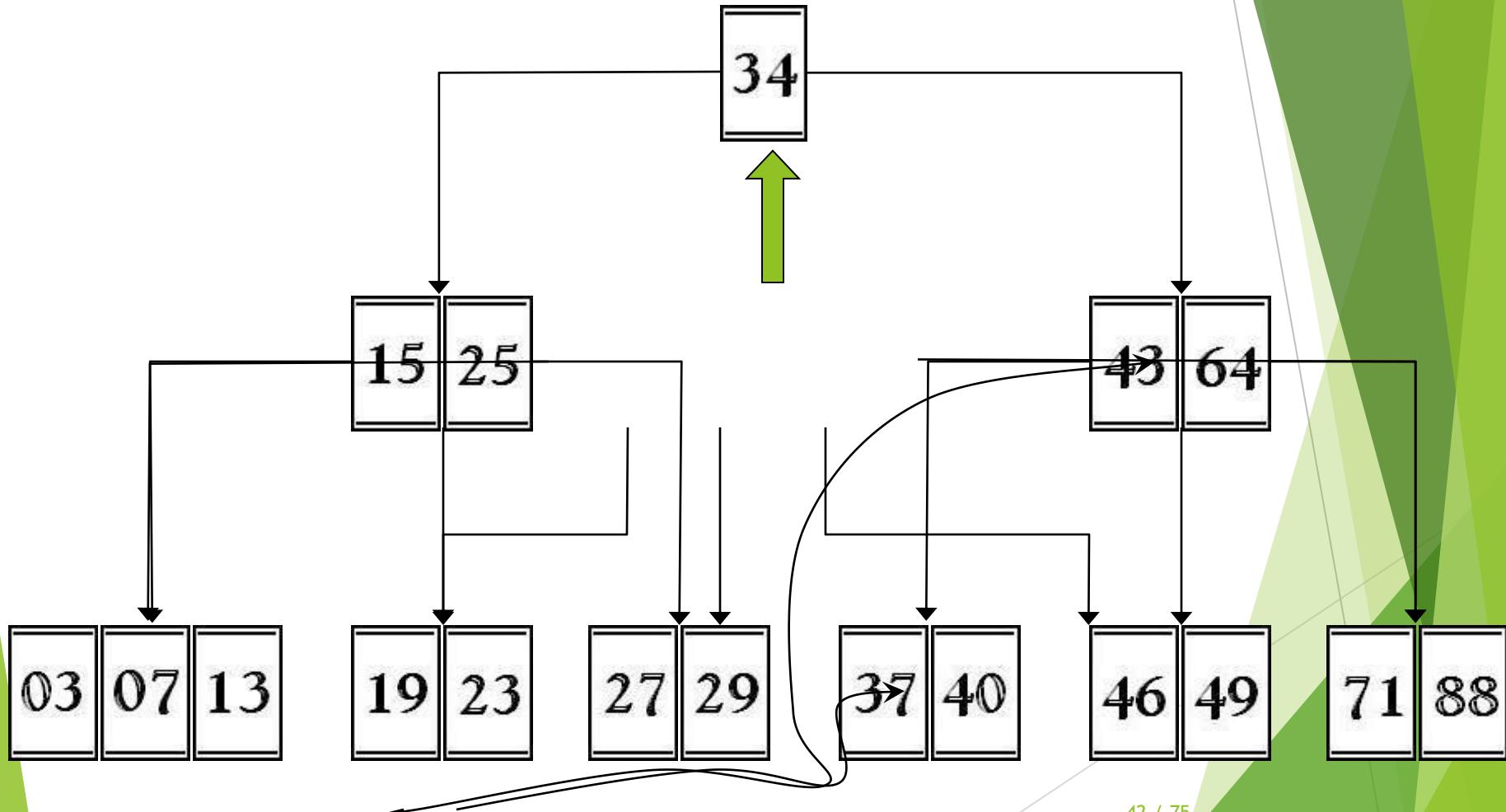
Ažuriranje datoteke s B-stablon

► Tehnika spajanja



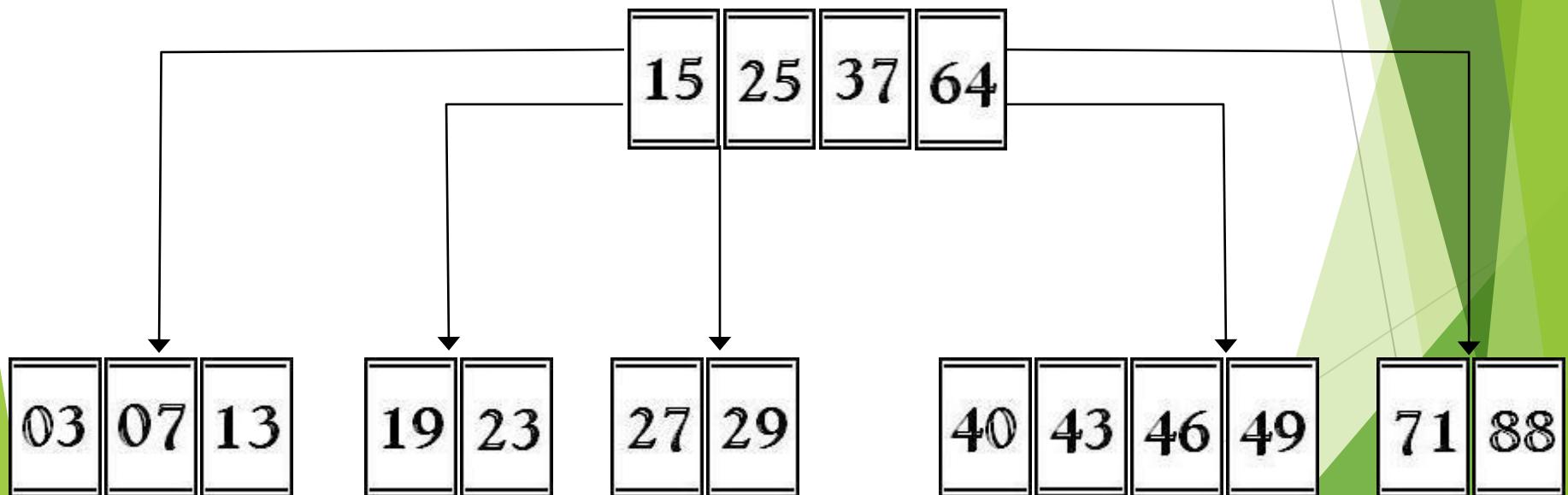
Ažuriranje datoteke s B-stablon

- ▶ Tehnika spajanja



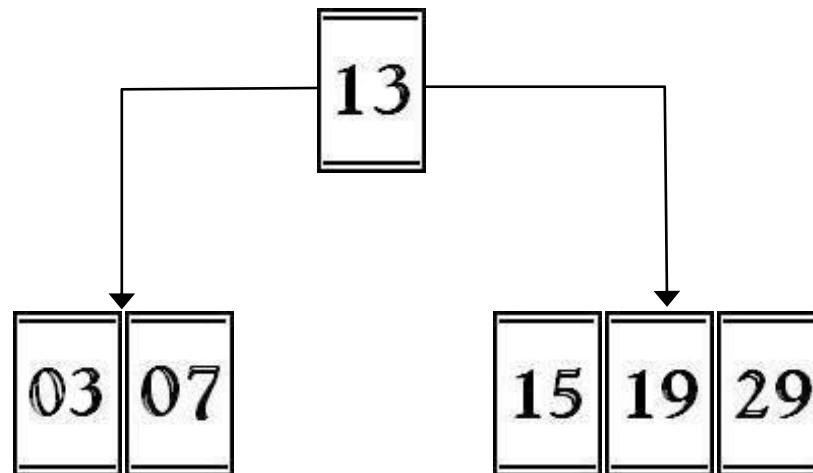
Ažuriranje datoteke s B-stablon

- ▶ Tehnika spajanja



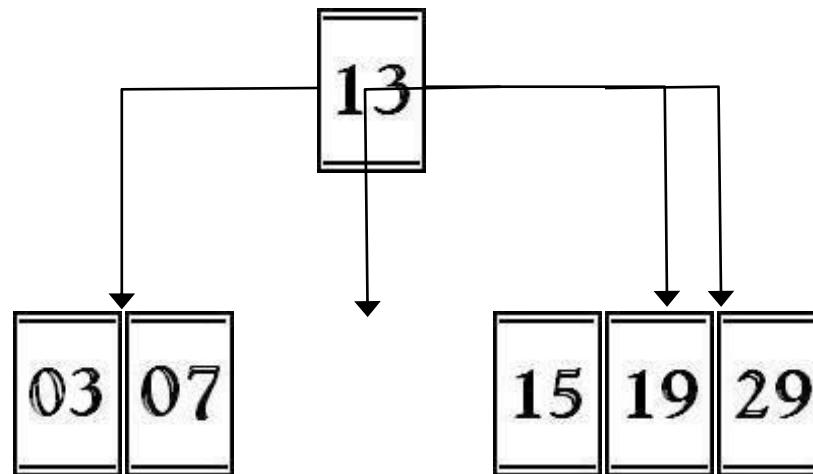
Ažuriranje datoteke s B-stablon

- ▶ Tehnika pozajmljivanja



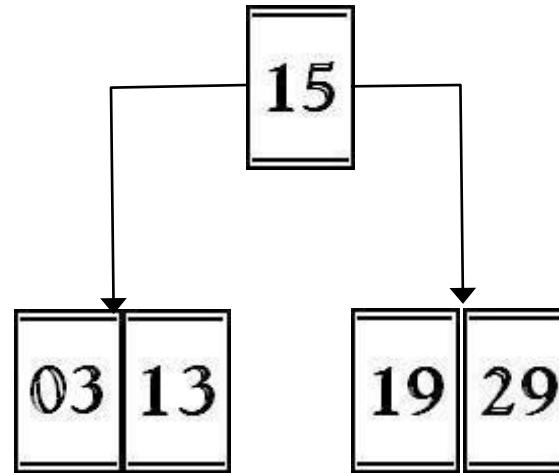
Ažuriranje datoteke s B-stabljom

- ▶ Tehnika pozajmljivanja



Ažuriranje datoteke s B-stablon

- ▶ Tehnika pozajmljivanja



Sadržaj

- ▶ Osnovno B-stablo
- ▶ Formiranje datoteke s B-stablotom
- ▶ Traženje u datoteci s B-stablotom
- ▶ Obrada datoteke s B-stablotom
- ▶ Ažuriranje datoteke s B-stablotom
- ▶ Ocena karakteristika datoteka s B-stablotom
- ▶ B*-stablo, B[#]-stablo, B⁺-stablo
- ▶ Indeks-sekvencijalna datoteka s B-stablotom

Ocena karakteristika datoteka s B-stabljom

► Prednosti

- ▶ pokazuju uravnotežene performanse pri direktnoj i redoslednoj obradi
- ▶ ne kvare se performanse obrade u vremenu
 - ▶ kao posledica naknadnih ažuriranja
- ▶ u odnosu na indeks-sekvencijalne datoteke
 - ▶ ne zahtevaju postojanje zone prekoračenja
 - ▶ struktura primarne zone odgovara strukturi serijske datoteke
 - ▶ maksimalni broj pristupa pri traženju slučajno odabranog sloga moguće je unapred proceniti
 - ▶ selektivnost indeksa po svakom pristupnom putu je ista

Ocena karakteristika datoteka s B-stablon

► Nedostaci

- nije najpogodnije rešenje za redoslednu obradu
 - broj pristupa datoteci zavisi od broja slogova vodeće datoteke
 - primena modifikovanog algoritma simetričnog postupka prolaska kroz stablo pri traženju adrese lokacije logički narednog sloga
 - potreba višestrukog pristupanja neterminalnim čvorovima
- deljenje čvorova favorizuje izgradnju polupraznog B-stabla
 - kada se formiranje datoteke vrši na osnovu sekvensijalne ulazne datoteke
 - svi listovi, osim krajnjeg desnog, uvek poluprazni
 - česta situacija u praksi, kada se vrednosti ključa generišu automatski, inkrementiranjem

Sadržaj

- ▶ Osnovno B-stablo
- ▶ Formiranje datoteke s B-stablotom
- ▶ Traženje u datoteci s B-stablotom
- ▶ Obrada datoteke s B-stablotom
- ▶ Ažuriranje datoteke s B-stablotom
- ▶ Ocena karakteristika datoteka s B-stablotom
- ▶ B*-stablo, B#-stablo, B+-stablo
- ▶ Indeks-sekvencijalna datoteka s B-stablotom

B*-stablo, B#-stablo, B+-stablo

► B*-stablo

- ▶ strukturalno isto, kao osnovno B-stablo
- ▶ **tehnika prelivanja**
 - ▶ ublažavanje problema favorizacije polupraznog stabla
 - ▶ primenjuje se pri upisu novog elementa u stablo
 - ▶ kada je čvor C_e u koji se upisuje element kompletan, a
 - ▶ bar jedan od njegovih susednih čvorova C_f sadrži $m_f < 2r$ elemenata
 - ▶ formira se u OM uređeni niz od $m_f + 2r + 1 + 1$ elemenata
 - ▶ svi elementi susednog čvora, jedan nadređeni i $2r + 1$ elemenata čvora u koji se upisuje element
 - ▶ $\lfloor (m_f + 2r + 2) / 2 \rfloor$ elemenata smešta se u levi susedni čvor
 - ▶ $\lfloor (m_f + 2r + 2) / 2 \rfloor + 1$ -vi element smešta se u nadređeni čvor
 - ▶ preostali elementi smeštaju se u desni susedni čvor
- ▶ favorizacija 75% popunjenoosti stabla

B*-stablo, B[#]-stablo, B⁺-stablo

- ▶ **B[#]-stablo**
 - ▶ varijanta B*-stabla
 - ▶ garantuje se minimalna popunjenošć 66%
 - ▶ za sve čvorove na svim nivoima hijerarhije, osim na prva dva
 - ▶ **tehnika distribuiranog deljenja**
 - ▶ primenjuje se pri upisu novog elementa u stablo
 - ▶ kada je čvor C_e u koji se upisuje element kompletan i kompletni su svi njegovi susedi
 - ▶ elementi susednih čvorova C_e i C_f distribuiraju se u tri čvora

B*-stablo, B#-stablo, B+-stablo

► B#-stablo

► tehnika distribuiranog deljenja

- ▶ elementi susednih čvorova C_e i C_f distribuiraju se u tri čvora
- ▶ formira se u OM uređeni niz od $2r + 2r + 1 + 1$ elemenata
 - ▶ $2r$ elemenata susednog čvora, jedan nadređeni i $2r + 1$ elemenata čvora u koji se upisuje element
 - ▶ $\lfloor (4r + 2) / 3 \rfloor$ elemenata smešta se u levi susedni čvor
 - ▶ $\lfloor (4r + 2) / 3 \rfloor + 1$ -vi element smešta se u nadređeni čvor
 - ▶ $\lfloor (4r + 1) / 3 \rfloor$ elemenata smešta se u srednji susedni čvor
 - ▶ $\lfloor (4r + 2) / 3 \rfloor + \lfloor (4r + 1) / 3 \rfloor + 2$ -gi element smešta se u nadređeni čvor
 - ▶ preostalih $\lfloor 4r / 3 \rfloor$ elemenata smešta se u novoalocirani, desni susedni čvor

B*-stablo, B#-stablo, B+-stablo

► **B⁺-stablo**

- modifikacija osnovnog B-stabla
 - prevazilaženje glavnog nedostatka osnovnog B-stabla, ili neke od njegovih varijanti
 - pri traženju logički narednog sloga potrebno je pristupati svim čvorovima stabla
- u slučaju B⁺-stabla, za traženje logički narednog sloga dovoljno je pristupati samo listovima

B*-stablo, B#-stablo, B+-stablo

► **B⁺-stablo**

- strukturalno drugačije od osnovnog B-stabla
 - vrednosti ključa svih slogova nalaze se uređene u listovima
 - svi listovi su dvostruko spregnuti
 - sadrže informaciju o svim logičkim vezama između slogova
 - čvorovi na višim nivoima stabla traženja sadrže najmanje vrednosti ključa iz svakog lista osim iz krajnjeg levog
 - vrednosti ključa čvorova koji ne predstavljaju listove, u nadređenim čvorovima ne ponavljaju se
 - logika formiranja neterminalnih čvorova odgovara strukturi osnovnog B-stabla

B*-stablo, B#-stablo, B+-stablo

► Format čvora B+-stabla

- niz elemenata
- svaki element: dvojka (k_e, P_e) , $e \in \{1, \dots, m\}$
 - k_e - vrednost ključa sloga S_i ($i \in \{1, 2, \dots, N\}$)
 - P_e - pokazivač ka podstablu, za neterminalni čvor ili bloku u primarnoj zoni, u slučaju lista

Zaglavljje bloka	P_0								Neiskorišćeni prostor
		k_1	P_1	k_2	P_2	\dots	k_m	P_m	

► Uslovi stabla traženja

- $(\forall i \in \{1, \dots, m - 1\})(k_i < k_{i+1})$
- $(\forall k \in K(P_0))(k < k_1)$
- $(\forall i \in \{1, \dots, m - 1\})(\forall k \in K(P_i))(k_i \leq k < k_{i+1})$
- $(\forall k \in K(P_m))(k_m \leq k)$

B^{*}-stablo, B[#]-stablo, B⁺-stablo

► Formiranje B⁺-stabla

- ▶ analogan postupak formiranju osnovnog B-stabla, ili neke njegove varijante (B^{*} ili B[#]-stabla)
 - ▶ mogu se koristiti postupci prelivanja, običnog ili distribuiranog deljenja, na uobičajen način
- ▶ **tehnika deljenja lista - specifičnosti**
 - ▶ formira se u OM uređeni niz od $2r + 1$ elemenata
 - ▶ svi elementi datog čvora i jedan novododati
 - ▶ $\lfloor (2r + 1) / 2 \rfloor = r$ elemenata smešta se u levi susedni čvor
 - ▶ $r + 1$ -vi element smešta se u nadređeni čvor
 - ▶ **$r + 1$ -vi element** i svih preostalih r elemenata smešta se u novoalocirani, desni susedni čvor
 - ▶ ponavljanje istog elementa i u nadređenom čvoru i u desnom susedu
 - ▶ neterminalni čvorovi dele se na isti način kao kod B-stabla

B*-stablo, B#-stablo, B+-stablo

► Formiranje B+-stabla

► tehnika prelivanja u listovima - specifičnosti

- ▶ formira se u OM uređeni niz od $m_f + 2r + 1 + 1$ elemenata
 - ▶ svi elementi susednog čvora, jedan nadređeni i $2r + 1$ elemenata čvora u koji se upisuje element
 - ▶ $\lfloor (m_f + 2r + 2) / 2 \rfloor$ elemenata smešta se u levi susedni čvor
 - ▶ $\lfloor (m_f + 2r + 2) / 2 \rfloor + 1$ -vi element smešta se u nadređeni čvor
 - ▶ $\lfloor (m_f + 2r + 2) / 2 \rfloor + 1$ -vi element i svi preostali elementi smeštaju se u desni susedni čvor
- ▶ prelivanje u neterminalnim čvorovima vrši se na isti način kao kod osnovnog B-stabla

B*-stablo, B#-stablo, B+-stablo

► Traženje u B+-stablu

► traženje logički narednog sloga

- vrši se isključivo u listovima B+-stabla
- primenom kombinacije metoda linearног traženja i metode praćenja pokazivača
- vrši se od tekućeg elementa B+-stabla

► redosledna obrada

- prolazak kroz sve listove B+-stabla i pristupanje primarnoj zoni

$$C_h^{min} + \left\lceil \frac{N_v^u}{f} \right\rceil \leq R_{uk}^r \leq C_h^{max} + N_v^u$$

$$\left\lceil \frac{N}{2r} \right\rceil + \left\lceil \frac{N_v^u}{f} \right\rceil \leq R_{uk}^r \leq \left\lfloor \frac{N}{r} \right\rfloor + N_v^u$$

B*-stablo, B[#]-stablo, B⁺-stablo

► Traženje u B⁺-stablu

► traženje slučajno odabranog sloga

► počinje u korenu B⁺-stabla i uvek se završava u jednom od listova - i za uspešno i za neuspešno traženje

► ako je $a < k_e$, dalje se prati pokazivač P_{e-1}

► ako je $a = k_e$, prati se pokazivač P_e - stiže se do lista

► ako je $a > k_m$, prati se pokazivač P_m

► broj pristupa, uspešno

► samo jedan bafer u OM za stablo pristupa

$$R_u = h^+ + 1$$

► broj pristupa, neuspešno

► samo jedan bafer u OM za stablo pristupa

$$R_n = h^+$$

B^{*}-stablo, B[#]-stablo, B⁺-stablo

► Traženje u B⁺-stablu

- visina kompletног B⁺-stabla

$$C_h^{kp} = (2r+1)^{h-1}, \quad N = 2rC_h^{kp} \Rightarrow \frac{N}{2r} = (2r+1)^{h-1}$$

$$h_{kp}^+ = 1 + \log_{2r+1} \frac{N}{2r}, \quad h_{min}^+ = 1 + \left\lceil \log_{2r+1} \frac{N}{2r} \right\rceil$$

- visina polupraznog B⁺-stabla

$$C_h^{pp} = 2(r+1)^{h-2}, \quad N = rC_h^{pp} \Rightarrow \frac{N}{2r} = (r+1)^{h-2}$$

$$h_{pp}^+ = 2 + \log_{r+1} \frac{N}{2r}, \quad h_{max}^+ = 2 + \left\lfloor \log_{r+1} \frac{N}{2r} \right\rfloor$$

B*-stablo, B[#]-stablo, B⁺-stablo

► Visine B-stabla i B⁺-stabla

► $h_{min}^+ \leq h^+ \leq h_{max}^+$

N	r	h_{min}	h_{max}	h_{min}^+	h_{max}^+
10^3	50	2	2	2	2
10^4	50	2	3	2	3
10^5	50	3	3	3	3
10^6	50	3	4	3	4

$$h_{min}^+ = 1 + \left\lceil \log_{2r+1} \frac{N}{2r} \right\rceil$$

$$h_{max}^+ = 2 + \left\lfloor \log_{r+1} \frac{N}{2r} \right\rfloor$$

$$h_{min} = \lceil \log_{2r+1}(N+1) \rceil$$

$$h_{max} = 1 + \left\lfloor \log_{r+1} \frac{N+1}{2} \right\rfloor$$

B*-stablo, B[#]-stablo, B⁺-stablo

► Ažuriranje B⁺-stabla

- ▶ upis novog elementa u B⁺-stablo vrši se na isti način kao pri formiraju B⁺-stabla
- ▶ brisanje
 - ▶ briše se i element u listu i, ako je ponovljen, element sa istom vrednošću ključa u nadređenom čvoru
- ▶ brisanje elementa iz čvora sa r elemenata
 - ▶ tehniku pozajmljivanja
 - ▶ tehniku spajanja čvorova

Sadržaj

- ▶ Osnovno B-stablo
- ▶ Formiranje datoteke s B-stablom
- ▶ Traženje u datoteci s B-stablom
- ▶ Obrada datoteke s B-stablom
- ▶ Ažuriranje datoteke s B-stablom
- ▶ Ocena karakteristika datoteka s B-stablom
- ▶ B^* -stablo, $B^{\#}$ -stablo, B^+ -stablo
- ▶ Indeks-sekvencijalna datoteka s B-stablom

Indeks-sekvencijalna dat. s B-stablon

► Indeks-sekvencijalna datoteka s B-stablon

- modifikacija B⁺-stabla
 - prevazilaženje glavnog nedostatka B⁺-stabla, ili neke od njegovih varijanti
 - za sva uspešna traženja sloga, potrebno je pristupati primarnoj zoni
- u slučaju indeks-sekvencijalne datoteke s B-stablon, uspešna traženja završavaju u listovima
- strukturalno drugačija od B⁺-stabla
 - u listovima B⁺-stabla ne memorišu se parovi (k_e, P_e), već kompletni slogovi
 - pogodna u situacijama kada kapacitet sloga nije preveliki
 - kako bi se očuvalo solidan rang stabla

Indeks-sekvencijalna dat. s B-stablon

- ▶ **Indeks-sekvencijalna datoteka s B-stablon**
 - ▶ strukturalno drugačija od B^+ -stabla
 - ▶ Listovi B^+ -stabla postaju blokovi primarne zone
- ▶ **Primarna zona**
 - ▶ spregnuta struktura
 - ▶ slogovi uređeni saglasno rastućim vrednostima ključa
- ▶ **Zona indeksa**
 - ▶ spregnuta struktura = neterminalni čvorovi B^+ -stabla
 - ▶ osnovno B-stablo ili neka od varijanti (B^* , $B^\#$)
 - ▶ retko popunjeni indeks
 - ▶ u zonu indeksa propagiraju se najmanje vrednosti ključa svakog bloka primarne zone, osim prvog bloka

Indeks-sekvencijalna dat. s B-stablon

► Indeks-sekvencijalna datoteka s B-stablon

- Primarna zona - faktor blokiranja

- $f = 2r$

- Primarna zona - broj blokova

- kompletno stablo

$$B_{min} = \left\lceil \frac{N}{2r} \right\rceil = \left\lceil \frac{N}{f} \right\rceil$$

- poluprazno stablo

$$B_{max} = \left\lceil \frac{N}{r} \right\rceil = \left\lceil \frac{2N}{f} \right\rceil$$

- broj blokova kreće se u rasponu

$$B_{min} \leq B \leq B_{max}$$

Indeks-sekvencijalna dat. s B-stablim

► **Formiranje**

- ▶ na uobičajen način
 - ▶ može se poći od ulazne serijske ili ulazne sekvencijalne datoteke

► **Obrada**

- ▶ može se koristiti
 - ▶ i kao vodeća i kao obrađivana
 - ▶ i u režimu redosledne i u režimu direktne obrade
 - ▶ pogodnija za primenu u redoslednoj obradi
- ▶ u redoslednoj obradi, pristupa se (svim) blokovima primarne zone
 - ▶ pogodno da broj blokova primarne zone bude manji
 - ▶ za manje kapacitete slogova, postiže se veći r i manji B

Indeks-sekvencijalna dat. s B-stablon

► Traženje i ažuriranje

- postupci analogni onima koji se primenjuju u slučaju B⁺-stabla
- **traženje logički narednog sloga**
 - vrši se isključivo u blokovima primarne zone
 - primenom kombinacije metoda linearog traženja i metode praćenja pokazivača
 - vrši se od tekućeg sloga primarne zone

Indeks-sekvencijalna dat. s B-stablon

► Traženje i ažuriranje

► traženje slučajno odabranog sloga

- počinje u korenu B-stabla i uvek se završava u bloku primarne zone - i za uspešno i za neuspešno traženje

► broj pristupa, uspešno

- samo jedan bafer u OM za stablo pristupa

$$R_u = h^{is} + 1$$

► broj pristupa, neuspešno

- samo jedan bafer u OM za stablo pristupa

$$R_n = h^{is} + 1$$

Indeks-sekvencijalna dat. s B-stablon

► Traženje slučajno odabranog sloga

- visina kompletног B-stabla

$$C_{kp} = \frac{(2r+1)^h - 1}{2r}, \quad B - 1 = 2rC_{kp} \Rightarrow B = (2r+1)^h$$

$$h_{kp}^{is} = \log_{2r+1} B = \log_{2r+1} \frac{N}{2r}, \quad h_{min}^{is} = \left\lceil \log_{2r+1} \frac{N}{2r} \right\rceil$$

- visina polupraznog B-stabla

$$C_{pp} = 1 + 2 \frac{(r+1)^{h-1} - 1}{r}, \quad B - 1 = 1 + r(C_{pp} - 1) \Rightarrow B = 2(r+1)^{h-1}$$

$$h_{pp}^{is} = 1 + \log_{r+1} \frac{B}{2} = 1 + \log_{r+1} \frac{N}{2r}, \quad h_{max}^{is} = 1 + \left\lceil \log_{r+1} \frac{N}{2r} \right\rceil$$

Indeks-sekvencijalna dat. s B-stablot

- Visine B-stabla, B^+ -stabla i B-stabla IS datot.

- $h_{min}^{is} \leq h^{is} \leq h_{max}^{is}$

N	r	h_{min}	h_{max}	h_{min}^+	h_{max}^+	h_{min}^{is}	h_{max}^{is}
10^3	50	2	2	2	2	1	1
10^4	50	2	3	2	3	1	2
10^5	50	3	3	3	3	2	2
10^6	50	3	4	3	4	2	3

$$h_{min}^{is} = \left\lceil \log_{2r+1} \frac{N}{2r} \right\rceil$$

$$h_{min}^+ = 1 + \left\lceil \log_{2r+1} \frac{N}{2r} \right\rceil$$

$$h_{max}^{is} = 1 + \left\lfloor \log_{r+1} \frac{N}{2r} \right\rfloor$$

$$h_{max}^+ = 2 + \left\lfloor \log_{r+1} \frac{N}{2r} \right\rfloor$$

Indeks-sekvencijalna dat. s B-stablon

- ▶ **Ocena karakteristika indeksnih datoteka s B-stabloma**
 - ▶ pogodne i za direktnu i za redoslednu obradu
 - ▶ relativno brz pristup slučajno odabranom slogu
 - ▶ ne prevelik broj pristupa u redoslednoj obradi
 - ▶ za efikasniju redoslednu obradu koriste se indeks-sekvencijalne datoteke s B-stablon
 - ▶ u slučajevima ne prevelikog kapaciteta sloga
 - ▶ kompromisno rešenje pri izboru fizičke strukture podataka
 - ▶ nije idealna organizacija ni za redoslednu ni za direktnu obradu, ali
 - ▶ organizacija, prisutna u svim savremenim SUBP, bez izuzetka

Sadržaj

- ▶ Osnovno B-stablo
- ▶ Formiranje datoteke s B-stablotom
- ▶ Traženje u datoteci s B-stablotom
- ▶ Obrada datoteke s B-stablotom
- ▶ Ažuriranje datoteke s B-stablotom
- ▶ Ocena karakteristika datoteka s B-stablotom
- ▶ B*-stablo, B[#]-stablo, B⁺-stablo
- ▶ Indeks-sekvencijalna datoteka s B-stablotom

Pitanja i komentari



Indeksne datoteke s B-stablima

B-stablo, formiranje, traženje, ažuriranje i obrada indeksne datoteke sa B-stablom