



Datoteke sa više ključeva

*multilista, invertovana datoteka,
upiti na interval vrednosti ključa*

Sadržaj

- Pretraživanje datoteke
- Metoda pristupa datoteka sa više ključeva
- Multilista
- Invertovana datoteka
- Oblasti primena datoteka sa više ključeva

Pretraživanje datoteke

- Fizička struktura sekvencijalne i spregnute datoteke
 - poseduje informaciju o vezama između slogova u logičkoj strukturi podataka datoteke
 - pogodne za traženje logički narednog sloga
- Fizička struktura rasute datoteke
 - poseduje informaciju o funkcionalnoj vezi između vrednosti ključa sloga i adrese lokacije u koju je slog (verovatno) smešten
 - pogodna za traženje slučajno odabranog sloga

Pretraživanje datoteke

- Indeksne datoteke poseduju obe opisane osobine
 - efikasno traženje slogova u datoteci na osnovu poznate vrednosti primarnog ključa

Pretraživanje datoteke

- Pretraživanje datoteke predstavlja algoritam AP , koji definiše preslikavanje:

$$AP : dom(X) \rightarrow \mathcal{P}(S)$$

- $dom(X)$ je skup vrednosti sekundarnog ključa X
 - $\mathcal{P}(S)$ partitivni skup skupa pojava S tipa sloga datoteke
- Vrednost sekundarnog ključa se naziva argumentom pretraživanja
 - iz skupa S izdvaja podskup S' slogova sa istom vrednošću

Pretraživanje datoteke

- Dva karakteristična slučaja pretraživanja datoteke
 - $\{A_1, \dots, A_n\}$ skup obeležja tipa sloga datoteke
 - X_1, \dots, X_p , $p \geq 1$ članovi skupa $\{A_1, \dots, A_n\}$, koji predstavljaju sekundarne ključeve tipa sloga
 - $x_i \in \text{dom}(X_i)$ za $i = 1, \dots, p$
 - na osnovu konjunkcije zadatih vrednosti sekundarnih ključeva X_1, \dots, X_l

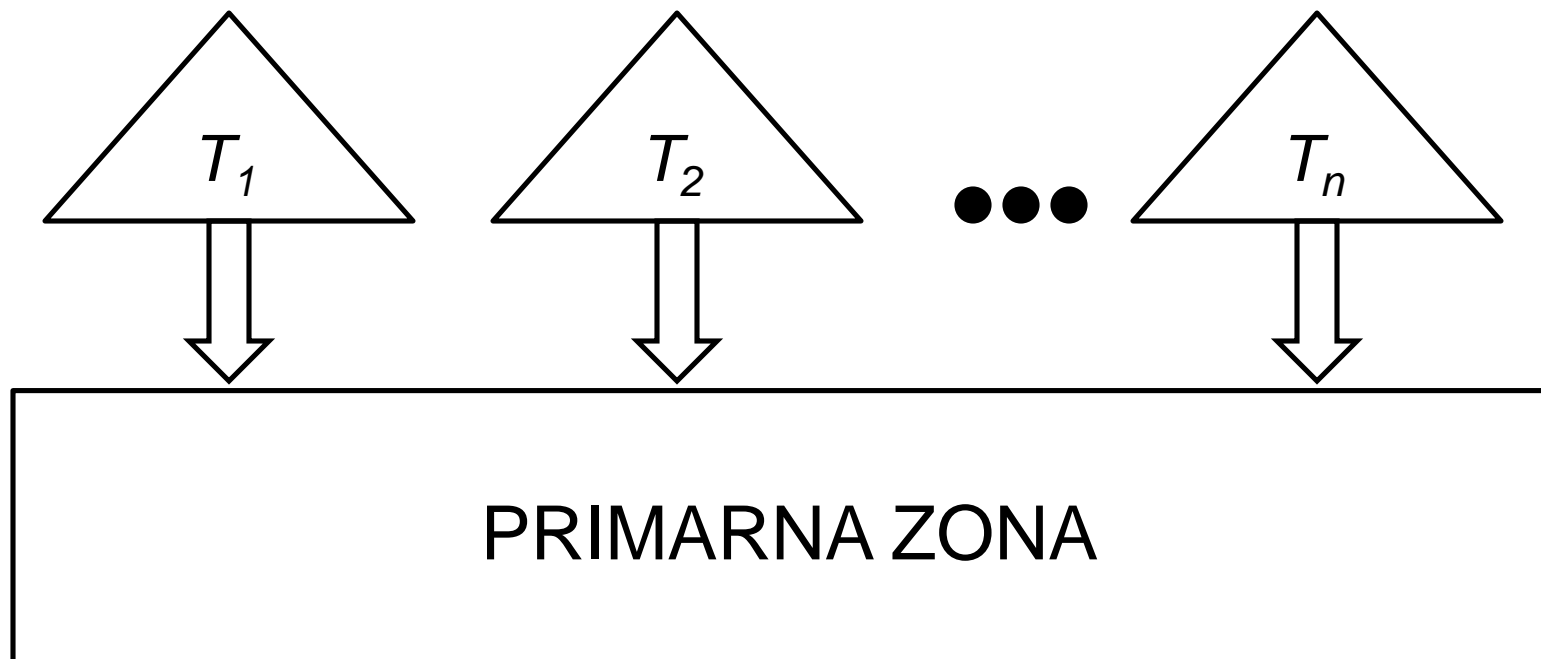
upit tipa: ? $(x_1 \wedge x_2 \wedge \dots \wedge x_l)$
 - na osnovu disjunkcije zadatih vrednosti sekundarnih ključeva X_1, \dots, X_m

upit tipa: ? $(x_1 \vee x_2 \vee \dots \vee x_m)$

Pretraživanje datoteke

- Datoteke sa više ključeva sadrže stabla traženja za svaki sekundarni ključ
 - stablo traženja je jedna od varijanti B-stabla
 - pretraživanje podrazumeva traženje adresa slogova
 - sa zadatim vrednostima sekundarnih ključeva u svakom od / stabala pristupa
 - /jednako broju sekundarnih ključeva navedenih u upitu
 - sa aspekta primarnog ključa, datoteka najčešće predstavlja indeksnu datoteku sa jednom od varijanti B-stabla

Pretraživanje datoteke



Pretraživanje datoteke

- Potreban broj pristupa R_k primarnoj zoni datoteke sa više ključeva,
 - pri pretraživanju konjunktivnog tipa, uzima celobrojne vrednosti iz intervala

$$[d(x_1 \wedge x_2 \wedge \dots \wedge x_l) / f, d(x_1 \wedge x_2 \wedge \dots \wedge x_l)]$$

- $d(x_1 \wedge x_2 \wedge \dots \wedge x_l)$ je kardinalni broj skupa S_k , podskupa skupa slogova datoteke

Pretraživanje datoteke

- Potreban broj pristupa R_d primarnoj zoni datoteke sa više ključeva,
 - pri pretraživanju disjunktivnog tipa, uzima celobrojne vrednosti iz intervala

$$[d(x_1 \vee x_2 \vee \dots \vee x_l) / f, d(x_1 \vee x_2 \vee \dots \vee x_l)]$$

- $d(x_1 \vee x_2 \vee \dots \vee x_l)$ je kardinalni broj skupa S_d , podskupa skupa slogova datoteke

Sadržaj

- Pretraživanje datoteke
- Metoda pristupa datoteka sa više ključeva
- Multilista
- Invertovana datoteka
- Oblasti primena datoteka sa više ključeva

Metoda pristupa datoteka sa više ključeva

- Indeksna metoda pristupa
 - operativni sistemi mainframe računara
 - poseduje metode pristupa za formiranje, korišćenje i ažuriranje indeksnih datoteka sa *B*-stablina
 - omogućeno definisanje alternativnih ključeva u programu za formiranje datoteke
 - programski jezici
 - korisnici sami pišu svoje metode
 - SUBP
 - poseduju sopstvene indeksne metode pristupa
 - koriste ih u izgradnji fizičkih struktura baza podataka
 - za izgradnju *B*-stabala sekundarnih ključeva potrebno je napraviti eksplicitan zahtev SUBP

Sadržaj

- Pretraživanje datoteke
- Metoda pristupa datoteka sa više ključeva
- Multilista
- Invertovana datoteka
- Oblasti primena datoteka sa više ključeva

Multilista

- Struktura primarne zone datoteke odgovara strukturi spregnuto organizovane datoteke, sa aspekta sekundarnih ključeva
 - svi slogovi sa istom vrednošću sekundarnog ključa obrazuju jedan lanac
 - broj lanaca odgovara zbiru različitih aktuelnih vrednosti tih obeležja – sekundarnih ključeva

Multilista

- Svakom lancu odgovara jedan indeks početka
 - sadrži uređenu trojku $(x_i, A_i, d(x_i))$
 - x_i vrednost sekundarnog ključa
 - A_i adresa prvog sloga u primarnoj zoni sa vrednošću sekundarnog ključa x_i
 - $d(x_i)$ dužina lanca slogova sa vrednošću sekundarnog ključa x_i

Multilista

- Indeks početka lanca
 - element stabla traženja sekundarnog ključa X_i
 - služi za pronalaženje adrese prvog sloga u lancu
 - slogova sa određenom vrednošću sekundarnog ključa
- Primarna zona multiliste, sadrži (dvostruko) spregnute (ciklične) lance slogova sa istom vrednošću sekundarnog ključa

Multilista

- Tip sloga datoteke je *VOZILO* (*REBR*, *PROI*, *BOJA*)
 - *REB* je registarski broj vozila
 - *PROI* je naziv proizvođača
 - *BOJA* je boja automobila
- dva sekundarna ključa
 - *PROI* i *BOJA*
 - $\text{dom}(\text{PROI}) = \{\text{Citroen}, \text{Folksvagen}, \text{Isuzu}, \text{Opel}, \text{Zastava}\}$
 - $\text{dom}(\text{BOJA}) = \{\text{bela}, \text{crvena}, \text{oker}, \text{plava}, \text{zelena}\}$

Multilista

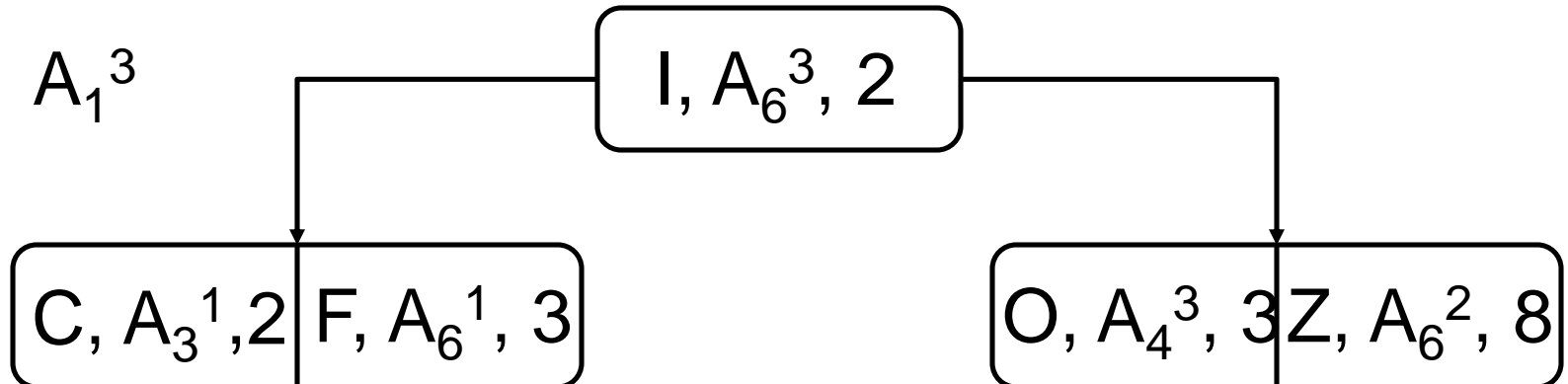
- Format lokacije za smeštaj sloga *VOZILO*

k	p	up	b	ub
----------	----------	-----------	----------	-----------

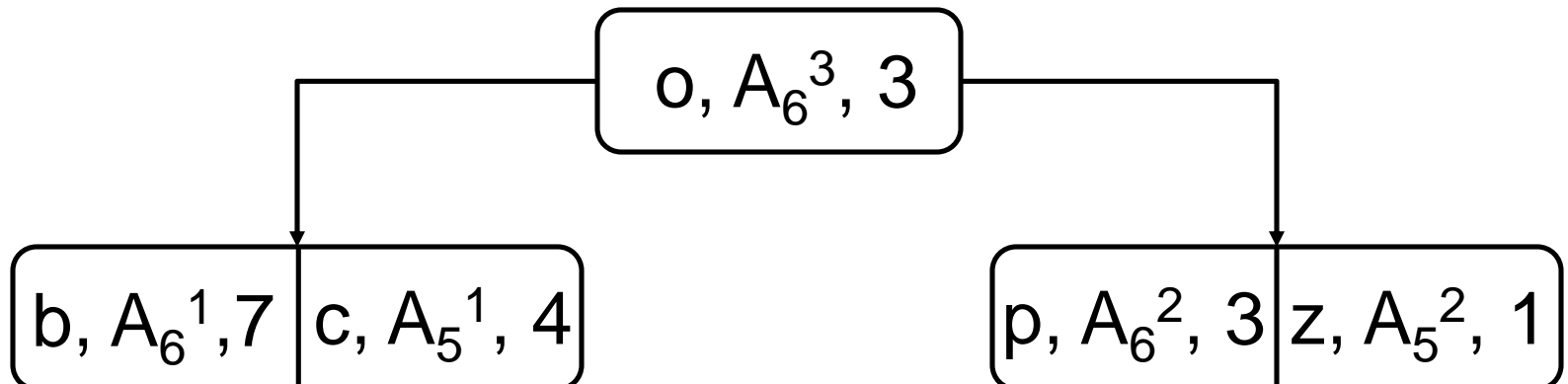
- *k* sadrži vrednost primarnog ključa *REBR*
- *p* sadrži vrednost sekundarnog ključa *PROI*
- *up* sadrži adresu lokacije narednog sloga sa istom vrednošću obeležja *PROI*
- *b* sadrži vrednost obeležja *BOJA*
- *ub* sadrži adresu lokacije narednog sloga sa istom vrednošću obeležja *BOJA*

Multilista

Zona indeksa sekundarnog ključa PROI



Zona indeksa sekundarnog ključa BOJA



Multilista

A_1	A_1^1					A_1^2					A_1^3				
	33 - 01	C	*	b	*	18 - 21	Z	*	c	*	71-13	O	*	p	*
A_2	A_2^1					A_2^2					A_2^3				
	01 - 03	O	A_1^3	b	A_1^1	13 - 19	Z	A_1^2	b	A_2^1	22 - 00	I	*	p	A_1^3
A_3	A_3^1					A_3^2					A_3^3				
	43 - 15	C	A_1^1	o	*	55 - 16	Z	A_2^2	b	A_2^2	55 - 17	Z	A_3^2	c	A_1^2
A_4	A_4^1					A_4^2					A_4^3				
	02 - 15	F	*	o	A_3^1	43 - 81	Z	A_3^3	c	A_3^3	13 - 68	O	A_2^1	b	A_3^2
A_5	A_5^1					A_5^2					A_5^3				
	33 - 14	Z	A_4^2	c	A_4^2	21 - 29	F	A_4^1	z	*	17 - 19	Z	A_5^1	b	A_4^3
A_6	A_6^1					A_6^2					A_6^3				
	19 - 00	F	A_5^2	b	A_5^3	43 - 83	Z	A_5^3	p	A_2^3	01 - 01	I	A_2^3	o	A_4^1

Multilista

- **Formiranje multiliste**
 - putem indeksne metode pristupa
 - u režimu redosledne ili režimu direktne obrade

Multilista

- Postupak formiranja datoteke
 - smeštanje slogova u sukcesivne lokacije primarne zone datoteke
 - istovremeno formiranje i stabla traženja primarnog i svih sekundarnih ključeva
 - ako slog sadrži novu vrednost sekundarnog ključa,
 - formira se novi element u odgovarajućem stablu traženja
 - ako slog ne sadrži novu vrednost sekundarnog ključa,
 - postojeći element se modifikuje
 - u element se upisuje adresa lokacije novog sloga i modifikuje se podatak o dužini lanca
 - adresa lokacije, do tada prvog sloga u lancu, upisuje se u polje pokazivača novog sloga

Multilista

- Pretraživanje multiliste
 - upit konjunktivnog tipa: ? $(x_1 \wedge x_2 \wedge \dots \wedge x_l)$
 - u stablu traženja svakog od l sekundarnih ključeva pronade se odgovarajući elemenat $(x_i, A_i, d(x_i))$
 - pristupanje slogovima najkraćeg lanca u primarnoj zoni
 - nakon pristupa slogu, proverava se da li zadovoljava kompletan uslov $(x_1 \wedge x_2 \wedge \dots \wedge x_l)$
 - nedostatak: pristupa se i slogovima koji ne zadovoljavaju kompletan uslov
 - broj pristupa može biti veći od $d(x_1 \wedge x_2 \wedge \dots \wedge x_l)$

Multilista

- upit disjunktivnog tipa: ? $(x_1 \vee x_2 \vee \dots \vee x_m)$
 - u stablu traženja svakog od sekundarnih ključeva pronade se odgovarajući elemenat $(x_i, A_i, d(x_i))$
 - pristupanje slogovima svih / lanaca
 - pretraživanje je uspešno ako postoji barem jedan elemenat $(x_i, A_i, d(x_i))$ u barem jednom od stabala traženja sekundarnog ključa, gde je $i \in \{1, \dots, m\}$
 - nedostatak: višestruko se pristupa slogovima koji zadovoljavaju uslov traženja po više od jednom literalu
 - broj pristupa može biti veći od $d(x_1 \vee x_2 \vee \dots \vee x_l)$

Multilista

- Ažuriranje multiliste
 - upis novog sloga isti je kao i u slučaju formiranja multiliste
 - kod brisanja se pronalaze i hronološki prethodni i naredni slogovi u svakom od lanaca
 - čiji elemenat je predstavljao brisani slog, da bi se ažurirali pokazivači
 - modifikacija vrednosti sekundarnog ključa u nekom slogu zahteva
 - da se taj slog isključi iz jednog i uključi u drugi lanac,
 - kao i da se na odgovarajući način ažurira stablo traženja odgovarajućeg sekundarnog ključa

Multilista

- Ocena karakteristika multiliste
 - Prednosti
 - pretraživanje po upitima konjunktivnog i disjunktivnog tipa
 - za broj sekundarnih ključeva $I = 1$, broj pristupa primarnoj zoni multiliste zadovoljava uslov minimalno potrebnog broja pristupa
 - stabla traženja sekundarnih ključeva sadrže manje elemenata nego stablo traženja primarnog ključa
 - Nedostaci
 - za broj sekundarnih ključeva $I > 1$, broj pristupa primarnoj zoni multiliste **ne zadovoljava** uslov minimalno potrebnog broja pristupa
 - povećanje potrebnog prostora za memorisanje sloga
 - generisanje unije podskupova slogova dobijenih pri realizaciji odgovora na upite disjunktivnog tipa

Sadržaj

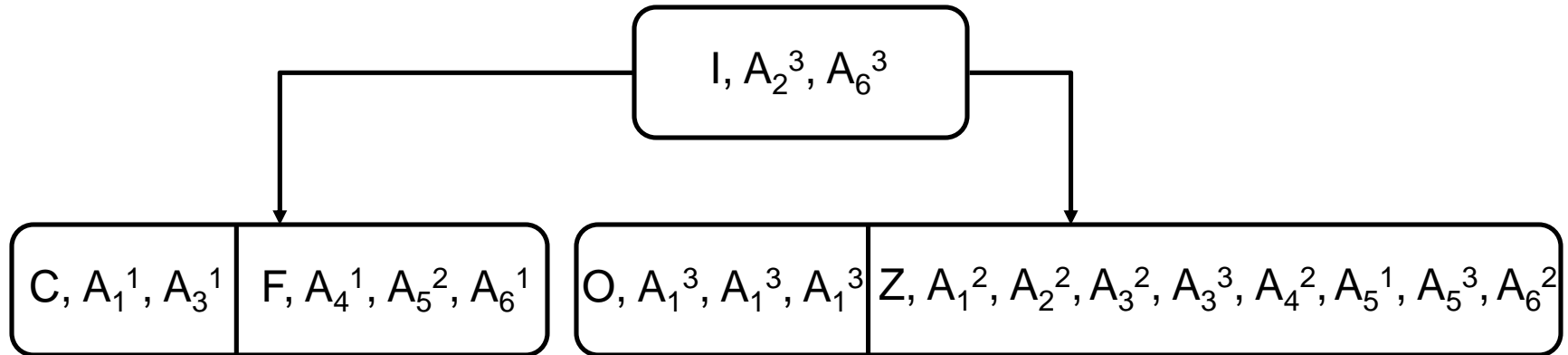
- Pretraživanje datoteke
- Metoda pristupa datoteka sa više ključeva
- Multilista
- Invertovana datoteka
- Oblasti primena datoteka sa više ključeva

Invertovana datoteka

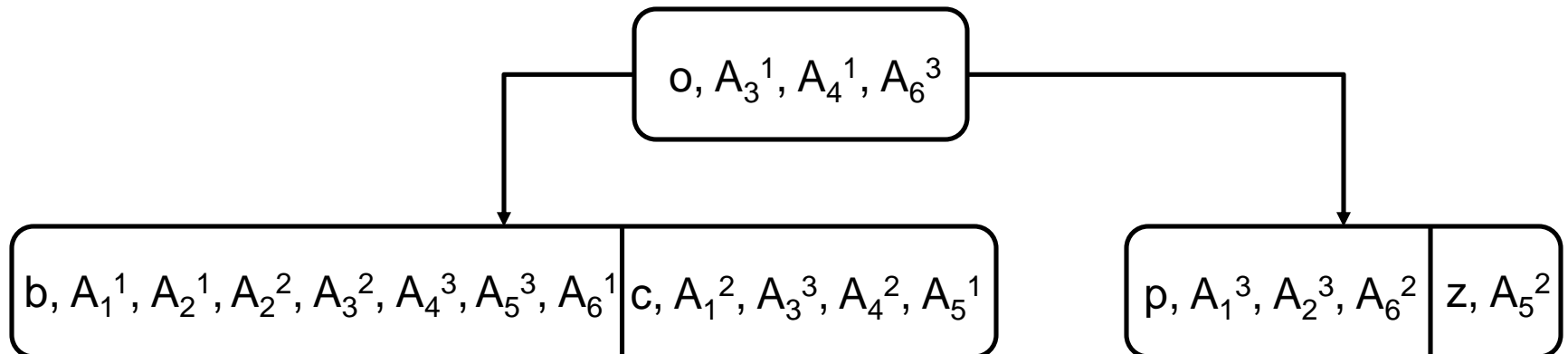
- Modifikovana struktura multiliste,
 - da bi se stvorili uslovi za efikasnije pretraživanje po upitima konjunktivnog i disjunktivnog tipa
- Invertovana organizacija datoteke nastaje tako što se ukone pokazivači iz primarne zone multiliste i
 - priključe adresi prvog sloga u odgovarajućim elementima stabala traženja sekundarnih ključeva

Invertovana datoteka

Zona indeksa sekundarnog ključa PROI



Zona indeksa sekundarnog ključa BOJA



Invertovana datoteka

A_1	A_1^1			A_1^2			A_1^3		
	33 - 01	C	b	18 - 21	Z	c	71-13	O	p

A_2	A_2^1			A_2^2			A_2^3		
	01 - 03	O	b	13 - 19	Z	b	22 - 00	I	p

A_3	A_3^1			A_3^2			A_3^3		
	43 - 15	C	o	55 - 16	Z	b	55 - 17	Z	c

A_4	A_4^1			A_4^2			A_4^3		
	02 - 15	F	o	43 - 81	Z	c	13 - 68	O	b

A_5	A_5^1			A_5^2			A_5^3		
	33 - 14	Z	c	21 - 29	F	z	17 - 19	Z	b

A_6	A_6^1			A_6^2			A_6^3		
	19 - 00	F	b	43 - 83	Z	p	01 - 01	I	o

Invertovana datoteka

- Formiranje invertovane datoteke
 - slično postupku formiranja multiliste
 - u primarnoj zoni se ne formiraju lanci slogova sa istim vrednostima sekundarnih ključeva
 - u elemente stabala traženja sekundarnih ključeva se upisuju adrese lokacija slogova sa istim vrednostima sekundarnih ključeva

Invertovana datoteka

- Pretraživanje invertovane datoteke
 - na osnovu vrednosti jednog sekundarnog ključa
 - izvršava se u invertovanoj datoteci sa istom efikasnošću kao i u multilisti
 - na osnovu vrednosti više sekundarnih ključeva
 - bolja efikasnost u odnosu na multilistu

Invertovana datoteka

- upit konjunktivnog tipa ? ($x_1 \wedge x_2 \wedge \dots \wedge x_l$)
 - u odgovarajućim stablima, pronalaze se elementi sa zadatim vrednostima sekundarnih ključeva
 - uslove upita zadovoljavaju samo slogovi čije adrese se nalaze u preseku svih / nizova adresa
 - traženje preseka adresa odvija se u operativnoj memoriji putem algoritama
 - pronađena adresa koristi se za pristup primarnoj zoni
 - broj pristupa R_k^i primarnoj zoni invertovane datoteke pri uspešnom pretraživanju konjunktivnog tipa, uzima celobrojne vrednosti iz intervala

$$[d(x_1 \wedge x_2 \wedge \dots \wedge x_l) / f, d(x_1 \wedge x_2 \wedge \dots \wedge x_l)]$$

Invertovana datoteka

- upit disjunktivnog tipa ? ($x_1 \vee x_2 \vee \dots \vee x_l$)
 - u operativnu memoriju učitava se l nizova adresa i upoređuju adrese iz svakog od l nizova
 - uvek se pristupa primarnoj zoni na adresi sa najmanjom vrednošću iz svih l nizova
 - kada se dođe do kraja jednog niza, nastavlja se upoređivanje adresa iz preostalih $l - 1$ nizova
 - pristup primarnoj zoni završava se kad se obrade svi elementi svih l nizova
 - broj pristupa R_d^j primarnoj zoni invertovane datoteke pri uspešnom pretraživanju disjunktivnog tipa, uzima celobrojne vrednosti iz intervala

$$[d(x_1 \vee x_2 \vee \dots \vee x_l) / f, d(x_1 \vee x_2 \vee \dots \vee x_l)]$$

Invertovana datoteka

- Ažuriranje invertovane datoteke
 - upis novog sloga u invertovanu datoteku vrši se kao i pri njenom formiranju
 - ažuriraju se stabla traženja svih ključeva datoteke
 - brisanje sloga iz primarne zone dovodi do brisanja adrese njegove lokacije iz stabala traženja svakog od sekundarnih ključeva
 - pomeranje svih adresa sa većom vrednošću, unutar elementa, za jedno polje ulevo
 - Modifikacija vrednosti sekundarnog ključa u slogu
 - dovodi do brisanja adrese njegove lokacije u jednom elementu i
 - njenog upisa u drugi elemenat stabla traženja istog sekundarnog ključa

Invertovana datoteka

- Ocena karakteristika invertovane datoteke
 - smanjen broj pristupa pri realizaciji konjunktivnih i disjunktivnih upita u odnosu na multilistu
 - promenljiva dužina čini nizove sa adresama lokacija veoma nepogodnim za smeštaj u elemente stabla traženja

Sadržaj

- Pretraživanje datoteke
- Metoda pristupa datoteka sa više ključeva
- Multilista
- Invertovana datoteka
- Oblasti primena datoteka sa više ključeva

Oblasti primena datoteka sa više ključeva



- Metode pristupa OS mainframe računara za izgradnju indeksnih datoteka
- Intenzivno se koriste u sistemima baza podataka
 - automatsko formiranje *B*-stabla za svaki primarni ključ
 - na osnovu deklarativnog zahteva korisnika, formiranje *B*-stabla sekundarnih ključeva
 - SQL dozvoljava definisanje svake kombinacije operacija konjunkcije, disjunkcije i negacije sekundarnih ključeva u upitima
 - ako stabla traženja ne postoje, pretraživanje se vrši metodom linearnog traženja
 - optimizator upita detektuje i koristi *B*-stabla sekundarnih ključeva automatski

Sadržaj

- Pretraživanje datoteke
- Metoda pristupa datoteka sa više ključeva
- Multilista
- Invertovana datoteka
- Oblasti primena datoteka sa više ključeva

Literatura

- Pavle Mogin: Strukture podataka i organizacija datoteka
 - Glava 15

Pitanja i komentari



Datoteke sa više ključeva

*multilista, invertovana datoteka,
upiti na interval vrednosti ključa*