

mieszczeniu głowic i wczytaniu indeksu ścieżek, odpowiedni element tego indeksu spowoduje włączenie odpowiedniej głowicy, w tym przypadku głowicy ścieżki drugiej, na której znajduje się szukany zapis. Szukanie zapisu z kluczem 741 prowadzi do tego samego elementu indeksu ścieżek, przy czym nadmiarowa skieruje poszukiwanie do listy (dwuzapisowej, w tym przypadku) zapisów nadmiarowych, związanych z daną ścieżką. Pierwszy element tej listy jest właśnie szukany zapisem.

Dołączanie - pierwszy element "pierwszego wiersza" indeksu cylindrów spowoduje wczytanie do PAO indeksu ścieżek pierwszego cylindra. Zapis z kluczem 79 będzie dołączany do ścieżki trzeciej. Ponieważ ścieżka jest już pełna, miejsce na wpisanie nowego zapisu można uzyskać "wypychając" zapis z kluczem 80 do lokalnego obszaru nadmiarów. Po operacji dołączenia pole klucza części podstawowej danego elementu indeksu zmieni zawartość z 80 na 79, zaś odsyłacz w części nadmiarowej będzie wskazywał na odpowiedni adres w obszarze nadmiarów, gdzie został zapisany "wypchnięty" zapis (Czytelnik proszony jest o zastanowienie się nad odpowiednimi zmianami na rys.10.8).

Zbiory indeksowo-sekwencyjne są jedynymi zbiorami zapewniającymi dostęp do informacji zarówno sekwencyjny, jak i bezpośredni. Dostęp sekwencyjny polega na sekwencyjnym przetwarzaniu każdej ścieżki obszaru podstawowego oraz po zlokalizowaniu początku listy zapisów nadmiarowych z nią związanych - na przejściu "wzdłuż listy", wykorzystując odsyłacze. Dostęp bezpośredni natomiast polega na każdorazowym (dla każdej wartości klucza) przejściu przez poszczególne poziomy indeksów, następnie sekwencyjnie przez ścieżkę podstawową oraz wg odsyłaczy przez listę nadmiarową, spełniając warunek: "klucz szukany  $\geq$  klucz w indeksie lub klucz na ścieżce".

### 10.3.1. Algorytm projektowania zbiorów indeksowo-sekwencyjnych

Rozważmy zbiór składający się z  $n_z$  zapisów o stałej długości  $l_z$ , przy czym długość klucza zapisu wynosi  $l_k$  [7], [8], [22].

### Algorytm ISD

1. Projektujemy obszar podstawowy wg kroku 1 i 2 algorytmu SPD (p.10.2.2). Musimy więc określić współczynnik blokowania  $n_{zb}$ , wielkość bloku  $l_b$ , całkowitą liczbę bloków  $n_b$  oraz liczbę bloków na ścieżce  $n_{bs}$  dla obszaru podstawowego. Należy uwzględnić, że wg standardów IBM bloki są zapisywane z kluczem \*).

2. Projektujemy lokalny obszar nadmiarów i globalny obszar nadmiarów, korzystając z informacji o przewidywanej liczbie  $n_{zn}$  zapisów, które mają być dodane do zbioru.

Dla nadmiaru lokalnego zarezerwowane jest kilka ścieżek w danym cylindrze, przy czym na ogół ich liczba  $n_{scl} \leq 2$ . Dla nadmiaru globalnego rezerwuje się kilka lub kilkanaście cylindrów, których liczba  $n_{cng}$  powinna być określona na podstawie typowych obliczeń projektowych wg algorytmu SPD, traktując zapisy dodawane jako pewien zbiór. Na ogół bloki w obszarach nadmiaru są jednozapisowe. Według standardów IBM są to bloki z kluczem, a poza tym podblok danych zawiera, oprócz danych użytkownika o długości  $l_z$ , także 10-znakowy odsyłacz, o którym była mowa powyżej. A więc na przykład całkowita długość bloku ostatniego  $L_{bon}$ , w obszarze nadmiaru będzie określona wzorem:

$$L_{bon} = l_{pd}'' + 10 + l_z + l_k .$$

3. Projektujemy indeks pierwszego poziomu, zwany dalej indeksem ścieżek, wg algorytmu SPD, traktując go jako zbiór składający się z  $n_{zi}$  zapisów, przy czym każdy zapis zawiera wartość klucza i adres zapisu lub grupy zapisów. Musimy więc określić liczbę bloków  $n_{bi}$ , tworzących indeks, liczbę bloków  $n_{bsi}$  w ścieżce oraz liczbę ścieżek indeksowych dla każdego cylindra wg wzoru:

$$n_{si} = \left\lceil \frac{n_{bi}}{n_{bsi}} \right\rceil .$$

\*) W podbloku klucza zapisywana jest wartość klucza "najstarszego" zapisu danego bloku (p.9.6.2.2).

Dla zbiorów wg standardów IBM należy przyjąć następujące założenia:

$$n_{zbi} = 1; l_{zi} = 10 \text{ zn}; n_{zi} = 2(n_{sc} - n_{scl}).$$

4. Projektujemy ewentualną nie zajętą część ostatniej ścieżki indeksu o długości  $l_{sr}$ , czyli dołączamy wolną część ścieżki do obszaru podstawowego.

Dla organizacji IBM słuszną jest następująca zależność:

$$l_{sr} = l_s - \text{mod}(n_{bi}, n_{bsi})^{*}) \cdot L_{bi}.$$

Jeśli  $l_{sr} < L_{bo}$ , na ostatniej ścieżce indeksu nie ma wolnego miejsca na dane użytkownika. W przeciwnym razie liczbę dodatkowych bloków użytkownika określamy wg znanego wzoru:

$$n_{br} = 1 + \left\lfloor (l_{sr} - L_{bo}) / L_b \right\rfloor.$$

5. Określamy liczbę bloków z danymi użytkownika zawartymi w cylindrze:

$$n_{bc} = n_{bs} \cdot [n_{sc} - (n_{si} + n_{scl})] + n_{br}.$$

6. Określamy liczbę cylindrów obszaru podstawowego:

$$n_{cp} = \left\lceil \frac{n_b}{n_{bc}} \right\rceil$$

oraz całkowitą liczbę cylindrów:

$$n_c = n_{cp} + n_{cng}.$$

7. Projektujemy indeks drugiego poziomu, zwany dalej indeksem cylindrów, traktując go jako zbiór składający się z  $n_{zic}$  zapisów, których struktura może być identyczna ze strukturą zapisów indeksu ścieżek.

---

\*) Zapis  $\text{mod}(x, y)$  oznacza resztę z dzielenia liczby  $x$  przez  $y$ .

Ponieważ dla każdego cylindra obszaru podstawowego mamy jeden zapis w indeksie cylindrów, więc  $n_{zic} = n_{cp}$ . Liczba ścieżek indeksu cylindrów jest określona wzorem:

$$n_{sic} = \left\lceil \frac{n_{zic}}{n_{bsi}} \right\rceil.$$

Jeśli indeks cylindrów zawiera zbyt wiele ścieżek, może być konieczne wprowadzenie indeksu trzeciego poziomu.

#### Przykład 10.6

Zbiór składający się z 20 000 zapisów 200-znakowych ma być zapisany jako zbiór indeksowo-sekwencyjny wg standardów IBM, przy czym 5000 zapisów użytkownik będzie dołączał sukcesywnie w terminie późniejszym. Długość klucza wynosi 12 znaków.

Dane:  $n_z = 15\ 000$ ;  $n_{zn} = 5000$ ;  $l_z = 200$  zn;  
 $l_k = 12$  zn.

Przyjmujemy:

$$n_{zb} = 3, \text{ czyli } l_b = 600 \text{ zn oraz } n_b = \left\lceil \frac{15000}{3} \right\rceil = 5000,$$

$$n_{scl} = 2.$$

Z założenia:

$$l_{zi} = l_{bi} = 10 \text{ zn}; \quad n_{zbi} = 1; \quad n_{zi} = 2 \cdot (10 - 2) = 16 = n_{bi},$$

$$n_{zbn} = 1, \text{ czyli } l_{bn} = 200 \text{ zn}; \quad n_{bn} = n_{zn} = 5000.$$

Liczba bloków na ścieżce w obszarze podstawowym i obszarze nadmiaru oraz całkowita liczba ścieżek nadmiaru:

$$n_{bs} = 1 + \left\lceil \frac{3625 - (20 + 12 + 600)}{81 + \lceil 1,049 \cdot (12 + 600) \rceil} \right\rceil = 5,$$

$$n_{bsn} = 1 + \left\lceil \frac{3625 - (20 + 12 + 10 + 200)}{81 + \lceil 1,049 \cdot (12 + 10 + 200) \rceil} \right\rceil = 11,$$

$$n_{sn} = \left\lceil \frac{5000}{11} \right\rceil = 455.$$

Liczba bloków na ścieżce indeksowej oraz liczba ścieżek indeksowych:

$$n_{bsi} = 1 + \left\lfloor \frac{3625 - (20 + 12 + 10)}{81 + \lceil 1,049 \cdot (12 + 10) \rceil} \right\rfloor = \left\lfloor \frac{3625 - 42}{105} \right\rfloor = 35,$$

$$n_{si} = \left\lceil \frac{16}{35} \right\rceil = 1.$$

Długość nie zajętej części ścieżki indeksowej oraz liczba bloków użytkownika w tej części:

$$l_{sr} = 3625 - 16 \cdot 105 = 1945 ,$$

$$n_{br} = 1 + \left\lfloor \frac{1945 - 632}{723} \right\rfloor = 2.$$

Liczba bloków użytkownika w cylindrze oraz liczba cylindrów obszaru podstawowego:

$$n_{bc} = 5 \cdot [10 - (1 + 2)] + 2 = 37,$$

$$n_{cp} = \left\lceil \frac{5000}{37} \right\rceil = 136.$$

Liczba cylindrów globalnego obszaru nadmiarowego oraz całkowita liczba cylindrów:

$$n_{cng} = \lceil (455 - 136 \cdot 2) / 10 \rceil = 19 ; n_c = 136 + 19 = 155.$$

Liczba ścieżek indeksu cylindrów:

$$n_{sic} = \left\lceil \frac{136}{35} \right\rceil = 4 .$$

### 10.3.2. Czasochłonność przetwarzania zbiorów indeksowo-sekwencyjnych

Dostęp do dowolnego zapisu w zbiorze indeksowo-sekwencyjnym, w celu np. jego aktualizacji, wymaga następujących operacji oraz następującego czasu trwania (przy założeniu, że zapis ten znajduje się w obszarze podstawowym):