

sprawdzić całą listę synonimów w celu stwierdzenia, czy nie jest podejmowana próba zapisania dwóch różnych zapisów z tym samym kluczem.

Ogólnie można powiedzieć, że dostęp do zbiorów o organizacji losowej wymaga mniej czasu niż do zbiorów o innej organizacji, jednak tracony jest pewien obszar pamięci. Organizacja ta jest stosowana wtedy, gdy czynnikiem krytycznym jest czas dostępu do informacji. Chcąc przyspieszyć wyszukiwanie grupy zapisów w zbiorze o organizacji losowej można wstępnie wygenerować adresy na podstawie wartości kluczy szukanych zapisów, a następnie transakcje wejściowe posortować wg wartości adresów. W ten sposób unikniemy zbędnych ruchów głowic w porównaniu z wyszukiwaniem zapisów osobno.

### 10.5. Zbiory inwersyjne

W dotychczasowych rozważaniach o organizacji zbiorów zajmowaliśmy się lokalizacją zapisu na podstawie wartości jego klucza pierwotnego, czyli klucza jednoznacznie identyfikującego dany zapis. Przedstawione powyżej organizacje zbiorów ułatwiały odpowiedzi na pytania i cechy (atrybuty\*) określonego obiektu, czyli o wartości kluczy wtórnych zapisu o danej wartości klucza pierwotnego. Na przykład dla zbioru KARTOTEKA OSOBOWA mogą to być pytania o zawód, staż pracy i zarobki pracownika o określonym numerze ewidencyjnym.

Czasami jednak, zwłaszcza w systemach wyszukiwania informacji, interesują nas pytania, dotyczące znalezienia obiektów o określonych cechach, czyli o określonych wartościach atrybutów. A więc zadanie polega na wyszukaniu zapisów (czyli pewnego podzbioru wartości klucza pierwotnego), których klucze wtórne przyjmują określone wartości.

\*) Klucze danego zapisu, opisującego pewien obiekt, nazywane są często (zwłaszcza w tzw. systemach wyszukiwania informacji) atrybutami tego obiektu lub atrybutami tego zapisu. Oczywiście atrybuty, podobnie jak klucze, przyjmują określone wartości, np. atrybut ZAWOD może przyjmować wartości: ELEKTRONIK, ELEKTRYK lub INFORMATYK, zaś atrybut WIEK - wartości 25, 36, 48 itp.

Wspomniane pytania należą zazwyczaj do następujących grup:

- pytania proste zawierające wartość pewnego atrybutu, np. ZAWÓD = ELEKTRYK, JĘZYK = ANGIELSKI,
- pytania o zakres zawierające określony zakres wartości pewnego atrybutu, np. ZAROBKI  $\leq$  4000, STAŻ PRACY  $>$  3.
- pytania złożone będące wyrażeniami logicznymi, utworzonymi z pytań prostych, pytań o zakres oraz operatorów logicznych: "I", "LUB", "NIE" itp., np. ((ZAWÓD = ELEKTRONIK) LUB (ZAWÓD = INFORMATYK)) I NIE (JĘZYK = ANGIELSKI).

Najprostszym sposobem znalezienia odpowiedzi na powyższe pytania dotyczące kluczy wtórnych jest przejrzanie całego zbioru i sprawdzenie każdego pola zapisu. Metoda ta, zbyt czasochłonna dla pojedynczych pytań, może być wystarczająco efektywna, jeśli pytania można zgrupować i wykonać tylko jeden raz przeszukiwanie sekwencyjne całego zbioru dla wszystkich pytań jednocześnie. Należy zwrócić uwagę, że jest to jedyna metoda, którą można stosować dla zbiorów w pamięciach o dostępie sekwencyjnym.

Jeśli jednak wymagane jest uzyskiwanie odpowiedzi w krótkim czasie, bezpośrednio po zadaniu pytania, oraz zbiory przeszukiwane są duże, konieczne jest zastosowanie pamięci o dostępie bezpośrednim oraz odpowiednich metod wyszukiwania zapisów. Jedną z najczęściej wykorzystywanych metod jest zastosowanie zbiorów inwersyjnych, zwanych też zbiorami odwrotnymi \*). Zbiorem inwersyjnym związanym z danym zbiorem, zorganizowanym wg klucza pierwotnego, nazywamy indeks, zawierający wartości pewnych atrybutów oraz listy zapisów, których atrybuty przyjmują wartości zawarte w indeksie (rys.10.10). Listy te nie zawierają pełnych zapisów, tylko ich adresy, przy czym mogą to być adresy symboliczne, względne i bezwzględne (p. 10.6). Atrybutami występującymi w indeksie mogą być dowolne

\*) Termin ten nie oznacza odwrócenia zbioru danych "do góry nogami", lecz po prostu zamianę roli zapisu i atrybutu. W zbiorach o organizacji dotychczas omówionej mieliśmy dany zapis i listę wartości jego atrybutów; w przypadku zbiorów inwersyjnych mamy listę zapisów o określonej wartości pewnego atrybutu.

klucze wtórne występujące w zapisach zbioru podstawowego, ich dowolne kombinacje uzyskane przez zastosowanie operatorów logicznych oraz atrybuty bezpośrednio nie występujące w zapisach (np. atrybuty przyspieszające uzyskiwanie odpowiedzi na pytania o zakres (rys.10.10)). Zastosowanie indeksów wprowadza pewną nadmiarową informację, dzięki której jednak można przyspieszyć wyszukiwanie zapisów na podstawie wartości ich kluczy wtórnych.

Zbiory inwersyjne można spotkać w życiu codziennym na ogół jednak pod innymi nazwami. Na przykład zbiorem odwrotnym w stosunku do słownika polsko-angielskiego jest słownik angielsko-polski. Podobnie skorowidz umieszczany często na końcu książek naukowych jest także zbiorem inwersyjnym związanym z całą książką.

a)

| Nr ewid. | Nazwisko<br>Imię | Zawód      | Staż<br>pracy | Języki |   |   | Płace |
|----------|------------------|------------|---------------|--------|---|---|-------|
|          |                  |            |               | R      | A | N |       |
| 1215     | BABACKI B.       | ENERGETYK  | 5             | 1      | 1 | 0 | 5500  |
| 1317     | DABACKI D.       | INFORMATYK | 2             | 1      | 1 | 1 | 4000  |
| 1415     | CABACKI C.       | ELEKTRONIK | 4             | 1      | 0 | 0 | 5000  |
| 1617     | ABACKI A.        | ELEKTRONIK | 3             | 0      | 1 | 0 | 4500  |
| 1805     | BABACKI C.       | ENERGETYK  | 7             | 0      | 1 | 1 | 6000  |

b)

| Atrybut    | Wartość atrybutu   | Lista wartości klucza pierwotnego                                      |
|------------|--|--|
| Zawód      | ELEKTRONIK<br>ENERGETYK<br>INFORMATYK                      | 1617, 1415<br>1215, 1805<br>1317                                       |
| Staż pracy | $\leq 4$<br>$> 3$  | 1317, 1415, 1617<br>1215, 1805, 1415                                   |
| Języki     | ROSYJSKI<br>ANGIELSKI<br>NIEMIECKI<br>ROSYJSKI I ANGIELSKI | 1215, 1317, 1415<br>1215, 1317, 1617, 1805<br>1317, 1805<br>1215, 1317 |
| Płace      | $\leq 4500$<br>$\geq 5000$                                 | 1317, 1617<br>1215, 1415, 1805   |

Rys.10.10. Struktura zbioru inwersyjnego: a) zbiór danych prosty, b) zbiór inwersyjny związany ze zbiorem prostym