

D.2. PRZETWARZANIE SEKWENCYJNE

W czasie trwania procesu przetwarzania danych przez maszynę cyfrową przepływają strumienie danych. (Terminu „dane“ używamy w nieco węższym sensie niż terminu informacje. Mianowicie, przez dane rozumiemy zwykle te informacje, które są przedmiotem przetwarzania). Strumienie te, wchodzące albo wychodzące poprzez różne urządzenia zewnętrzne maszyny, będziemy nazywali zbiorami danych. Zbiory danych przechodząc przez maszynę cyfrową ulegają przekształceniu na skutek wykonywania przez EMC programu — opisu w kodzie „zrozumiałym“ dla maszyny algorytmu przekształcania strumieni danych. Każdy taki zbiór danych jest podzielony na pewne fragmenty tworzące logiczną całość. Fragmenty takie będziemy nazywali rekordami (*record*). Koncepcja rekordów pochodzi od powszechnego nawyku biurokracji do używania typowych dokumentów, jak np.: zamówienia, dowody pobrania, dowody przyjęcia itp. Rekordy dzielą się na elementy, na których wykonywane są przesyłania i obliczenia lub które używane są do identyfikowania. Takie podstawowe elementy będziemy nazywali polami elementarnymi (*elementary field*). Czasami w obrębie rekordu wygodnie jest operować grupami pól elementarnych (*field*) łączących się pomiędzy sobą w sposób naturalny (porównaj p. D.10).

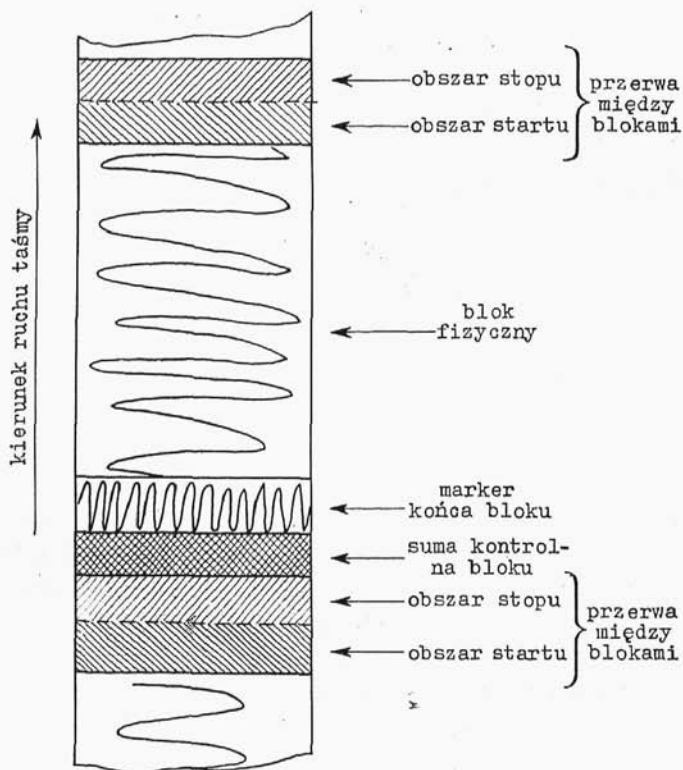
Jeden zbiór może się składać z jednego lub więcej typów rekordów. W przypadku gdy w zbiorze może występować więcej niż jeden typ rekordów, zachodzi konieczność zaopatrzenia każdego rekordu w informację identyfikującą typ rekordu. Oczywiście, że we wszystkich rekordach składających się na dany zbiór bez względu na typ, kod typu rekordu musi być zapisany w tym samym polu elementarnym. Przeważnie wyróżniamy jedno pole rekordu. Pole to przeznaczone jest do identyfikowania informacji zawartej w rekordzie i nazywane jest zwykle kluczem (*key*). Jeśli zbiór składa się np. z kart pracy (gdzie każda karta w tym przypadku stanowi oddzielny rekord), to kluczem może być numer robotnika. Należy wyjaśnić dwa fakty, które mogą budzić pewną wątpliwość: po pierwsze, jeden rekord może mieć więcej niż jeden klucz, po drugie, klucz nie musi być numeryczny, równie dobrze może być alfabetyczny czy też alfanumeryczny. W dalszym ciągu zbiory uporządkowane według jakiegoś klucza występującego we wszystkich rekordach składających się

na ten zbiór w kierunku malejących albo rosnących wartości tego klucza będziemy nazywali kartoteką.

Jeśli w procesie przetwarzania opracowujemy kolejne rekordy dwu (lub więcej) kartotek zgodnie względem siebie uporządkowanych, to taki proces przetwarzania będziemy nazywali przetwarzaniem sekwencyjnym.

W przypadku posługiwania się kartami perforowanymi jako nośnikiem informacji wymiary rekordu ograniczone są pojemnością karty (np. 80 znaków). Na ogół przyjmuje się, że jedna karta przeznaczona jest do przechowywania tylko jednego rekordu.

W przypadku posługiwania się drukarką liniową wymiary rekordu



Rys. 56. Schemat rozmieszczenia bloku fizycznego na taśmie magnetycznej

ograniczone są ilością znaków w linii. Na ogół przyjmuje się, że jedna linia przeznaczona jest na jeden rekord.

W przypadku taśm magnetycznych przyjmujemy zwykle nieco odmiennie konwencje. Wynika to z technicznych własności urządzeń taśmowych. Po pierwsze, taśma magnetyczna, w porównaniu z innymi urządzeniami pamięci zewnętrznymi, nie daje możliwości dokonywania jakichkolwiek zmian wcześniej zapisanych informacji bez przepisania pozostałych informacji na inną taśmę. Po drugie, taśma ma ograniczoną długość wynikającą z wymiarów szpuli, na którą jest nawijana. Po trzecie, dla dokonania zapisu lub odczytu taśma musi się poruszać z pewną określoną szybkością, natomiast w chwilach, kiedy nie dokonujemy ani zapisu, ani odczytu, taśma musi się zatrzymać. Dlatego też informacje na taśmie są zapisywane blokami fizycznymi o długości od kilku do kilkuset słów, tak jak to jest pokazane na rys. 56. Ze względu na przerwy pomiędzy blokami fizycznymi, w zależności od długości używanych bloków, na jednej taśmie można zapisać różne ilości informacji.

Wprowadzając następujące oznaczenia:

t — odcinek taśmy zużywany na zapisanie jednego słowa,

x — długość rekordu w słowach,

n — ilość rekordów w bloku fizycznym,

c — ilość słów pomocniczych, które muszą być zapisane w każdym bloku (obok rekordów) bez względu na użyteczną długość bloku liczoną w słowach,

g — średni odstęp między blokami, czyli łączna długość obszarów startu i stopu,

można podać wzór na średnią długość taśmy przypadającą na jeden rekord h , dla przypadku rekordów o stałej długości x ,

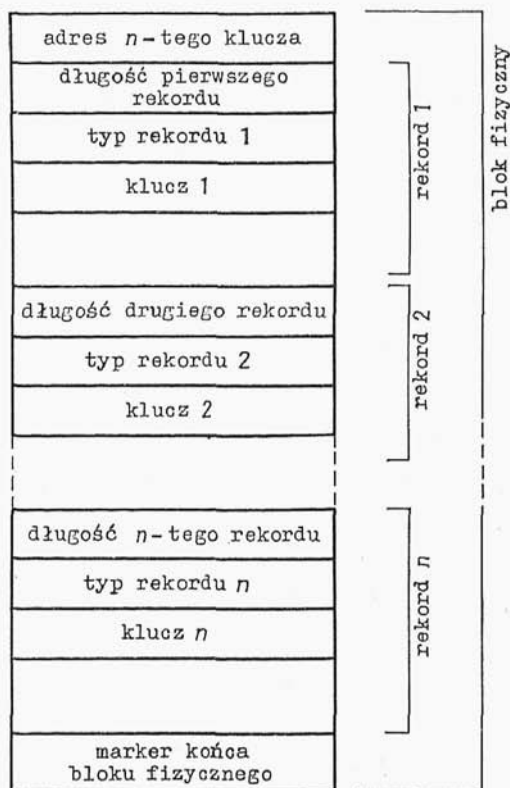
$$h = \frac{t(nx+c)+g}{n}. \quad (\text{D-1})$$

Dodatkowo oznaczając dostępną do zapisu długość taśmy przez L , można podać wzór na pojemność taśmy Q , liczoną w rekordach o stałej długości x ,

$$Q = \frac{nL}{t(nx+c)+g}. \quad (\text{D-2})$$

W pracy H. N. Ladena i T. R. Gildersleeve'a [26], obok wyżej podanych wzorów, można znaleźć szereg innych oszacowań.

Warto podkreślić w świetle dotychczasowych rozważań, że przy projektowaniu systemu elektronicznego przetwarzania danych stosowanie taśmy magnetycznej jako nośnika informacji pociąga za sobą koniecz-



Rys. 57. Struktura bloku fizycznego zawierającego n rekordów (bloków logicznych) o zmiennej długości uporządkowanych zgodnie z wartościami kluczy

ność minimalizowania ilości taśmo-przebiegów i maksymalizowania ilości zmian robionych w czasie generowania taśmy wyjściowej, tworzonej na podstawie jednej lub więcej taśm wejściowych. Czynnikiem decydującym

w tym zakresie jest podział informacji (na podstawie których operuje system) na zbiory. Przy projektowaniu systemu należy sprawę podziału na zbiory poddać bardzo wnikliwej analizie.

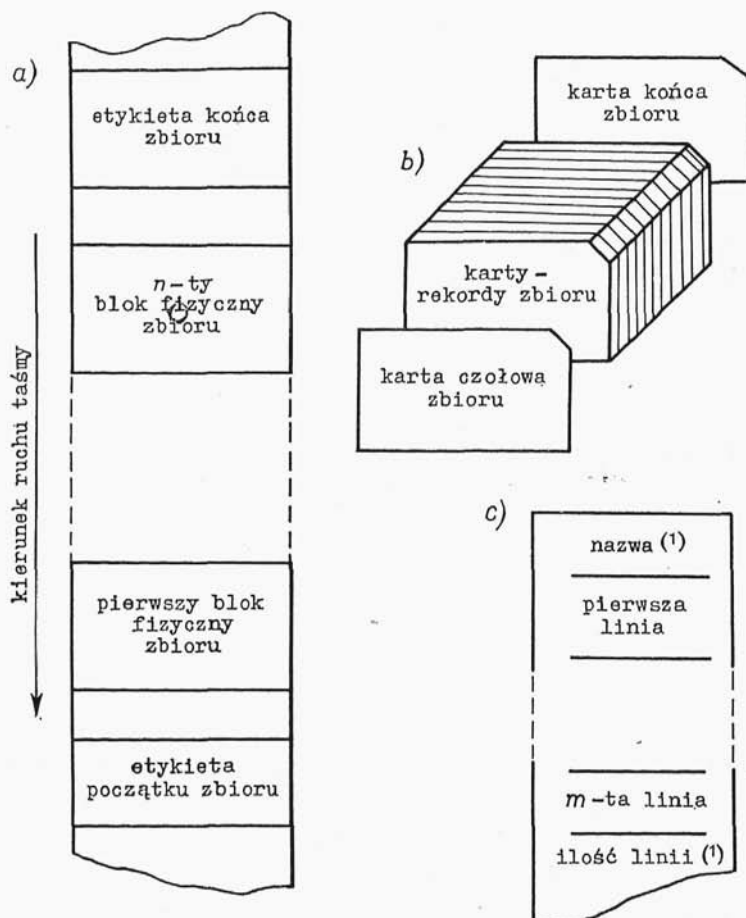
Dotychczas zajmowaliśmy się jedynie rekordami o ustalonej długości. Ze względu jednak na fizyczne własności taśmy znacznie lepiej można wyzyskać taśmę magnetyczną stosując rekordy o zmiennej długości. W tym przypadku każdy rekord należy zaopatrzyć w dodatkową informację — długość rekordu w słowach lub znakach. Bloki złożone z rekordów o zmiennej długości są, oczywiście, również zmiennej długości. W tej sytuacji należy określić maksymalną dopuszczalną długość bloku. Minimalna długość jest wówczas równa maksymalnej długości liczonej w słowach minus maksymalna długość rekordu liczona w słowach plus jeden. Na rysunku 57 pokazany jest przykład bloku fizycznego złożonego z rekordów o zmiennej długości.

Z dotychczasowych rozważań mogłoby wynikać, że można zwiększać długość bloku fizycznego nieograniczenie i że w ten sposób mamy coraz lepsze wyzyskanie taśmy magnetycznej. Jednakże, istnieje jeszcze drugi czynnik decydujący o długości bloków fizycznych. Na to aby zapisać czy też odczytać blok z taśmy magnetycznej, konieczne jest zarezerwowanie w pamięci operacyjnej centralnej jednostki EMC tzw. obszaru zbioru o wymiarach równych maksymalnej długości bloku fizycznego danego zbioru. W szczególnych przypadkach można przyporządkować dwu różnym zbiorom wspólnego obszaru zbioru w pamięci operacyjnej. Często jednak zachodzi konieczność zarezerwowania dwu lub więcej alternatywnych obszarów zbioru dla jednego zbioru. Jest to podyktowane stosunkowo małą szybkością zapisu i odczytu taśm magnetycznych w stosunku do szybkości centralnej jednostki EMC. Dlatego też długość bloków fizycznych jest kompromisem pomiędzy dążeniem do zwiększenia pojemności szpuli taśmy magnetycznej a dążeniem do przeznaczenia możliwie jak najmniejszej części pamięci operacyjnej na obszary zbiorów.

Zbiór przechowywany na taśmie magnetycznej poza rekordami informacyjnymi musi być zaopatrzony w tzw. etykiety. Rekordy zbioru są poprzedzone etykietą zbioru zawierającą nazwę zbioru, datę zapisania taśmy, numer generacji zbioru i datę przedawnienia, tj. datę, po której taśma może być zmaszana. Na końcu zbioru znajdują się tzw. etykiety

końca zbioru (rys. 58). W przypadku zbioru na kartach zamiast etykiety początku mamy kartę czołową, a zamiast etykiety końca zbioru — standardową kartę końca.

W przetwarzaniu sekwencyjnym kolejne opracowywanie rekordów na-



Rys. 58. Struktury zbiorów na różnych nośnikach

a) struktura zbioru o n blokach fizycznych przechowywanego na taśmie magnetycznej, b) struktura zbioru przechowywanego na kartach perforowanych, c) struktura zbioru — zestawienia mającego m rekordów (linii) — wydrukowanego za pomocą drukarki.

⁽¹⁾ te rekordy pomocnicze nie zawsze występują.

leżących do jednego lub więcej zbiorów nazywamy przebiegiem (*run, pass*). Każdy cykl przetwarzania składa się z szeregu przebiegów. W dalszym ciągu omówimy pewne typowe przebiegi, takie, jakie stosunkowo często występują w procesach przetwarzania sekwencyjnego.

Przebieg perforacji, redagowania i załadowywania informacji (*editing and converting*). Jak już wspominaliśmy, przy przetwarzaniu sekwencyjnym mamy do czynienia z partiami (*batch*)



Rys. 59. Dziurkarka kart 80-kolumnowych produkcji I.C.T. (fot. S. Porowski)

dokumentów pierwotnych. Dla każdej partii oblicza się sumę kontrolną i zapisuje się w książce kontrolnej obok numeru partii ilości dokumentów i daty otrzymania. Po perforacji (rys. 59) i sprawdzeniu (rys. 60) karty (lub taśmy perforowanej) odpowiedniki dokumentów źródłowych są wprowadzone do EMC. Zawartości formalnie poprawne kart zostają umieszczone na taśmie magnetycznej. Równocześnie drukowany jest raport błędów i sum kontrolnych. Na podstawie raportu błędów można wyselekcjonować karty formalnie błędne i poprawić je. W raporcie błę-

dów sumy kontrolne podawane są dla kart poprawnych, błędnych i łącznie dla obu rodzajów. Obok sum kontrolnych drukowane są również ilości kart. Liczby te są porównywane z książką kontrolną. W przypadku rozbieżności należy znaleźć przyczynę. Karty formalnie błędne po popra-

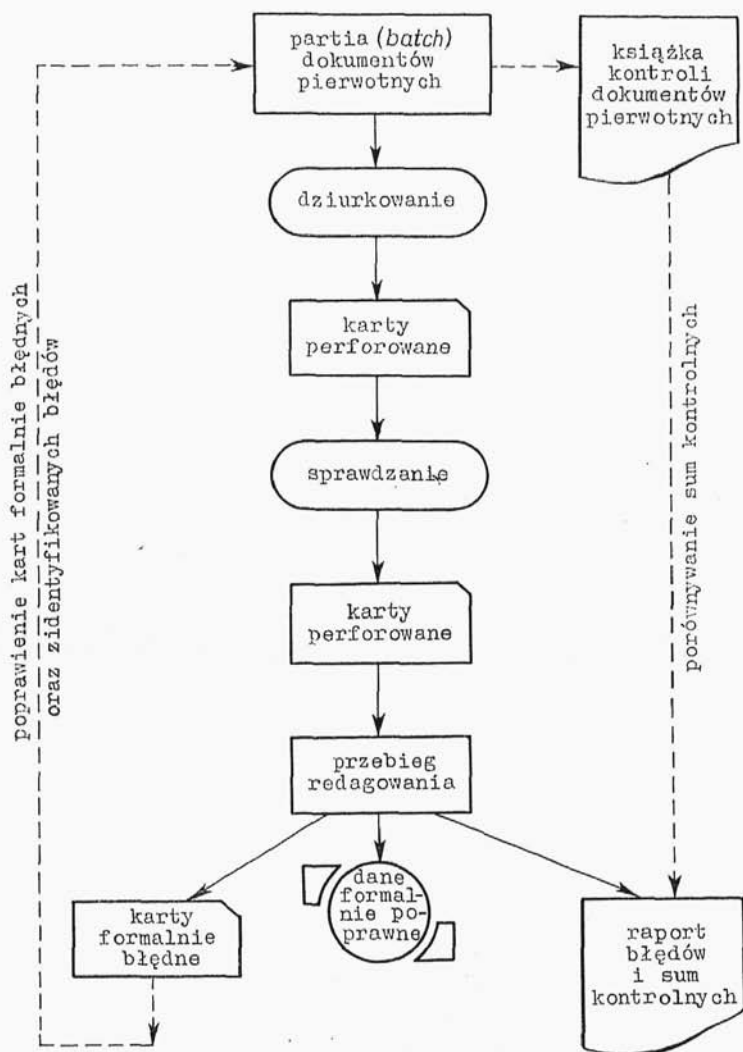


Rys. 60. Sprawdzarka kart 80-kolumnowych produkcji I.C.T. (fot. S. Porowski)

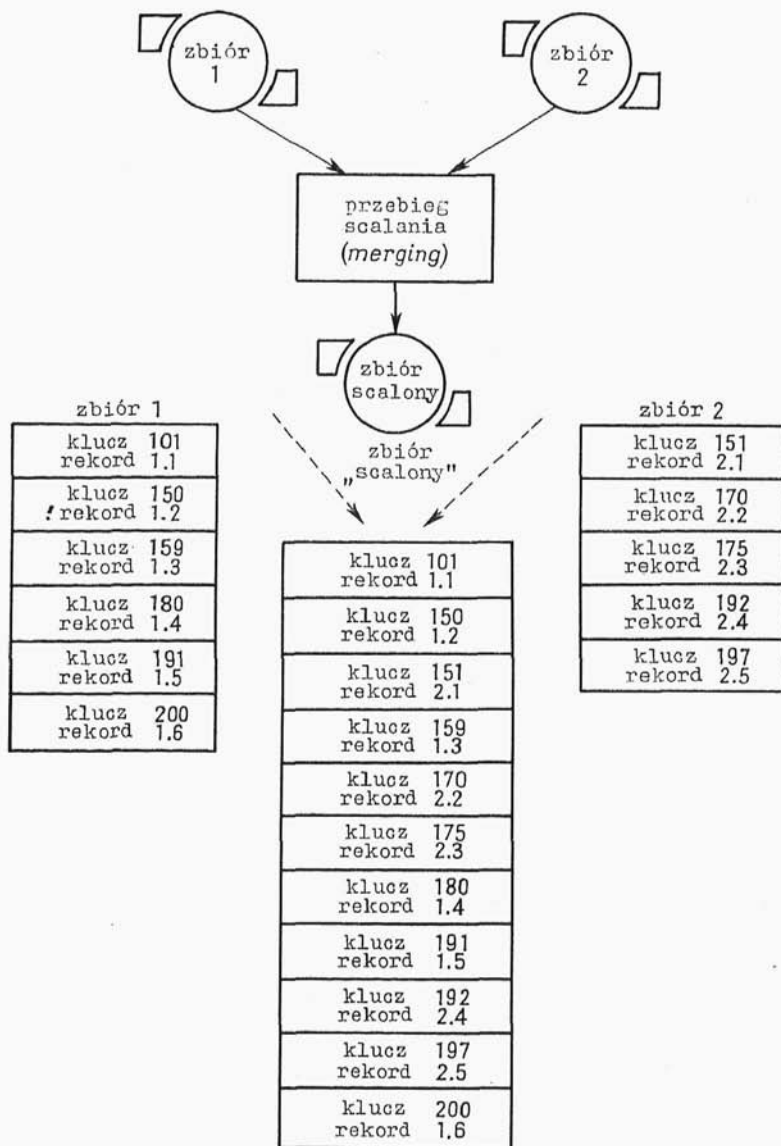
wieniu są ponownie wprowadzane do EMC. Cały proces powtarza się bądź do chwili usunięcia wszystkich błędów formalnych, bądź do chwili, w której przebieg załadowywania musi być zakończony. W tym ostatnim przypadku dokumenty formalnie błędne muszą być przetwarzane ręcznie i w odpowiedniej chwili włączone do właściwych zbiorów. Na rysunku 61 pokazany jest schemat takiego przebiegu.

Przebieg „scalania” (*merging*) służy do połączenia dwu lub więcej zbiorów zgodnie uporządkowanych (czyli kartotek) w jeden zbiór uporządkowany. Na rysunku 62 pokazano przykład scalania dwu zbiorów.

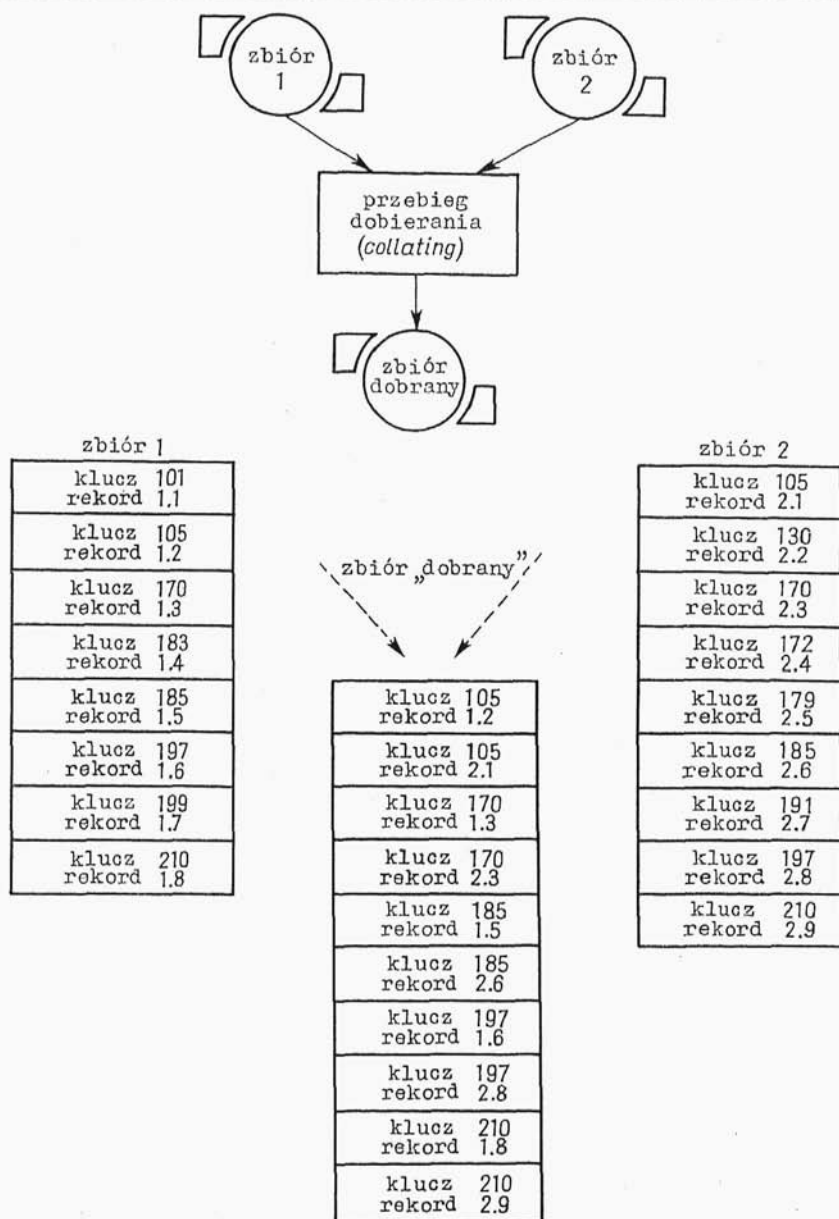
Przebieg „dobierania” (*collating*) służy do wybierania z dwu lub więcej zbiorów zgodnie uporządkowanych (czyli kartotek) wspólnego,



Rys. 61. Przebieg redagowania i załadowania (editing and converting) zawartości formalnie poprawnych kart perforowanych — odpowiedników dokumentów pierwotnych



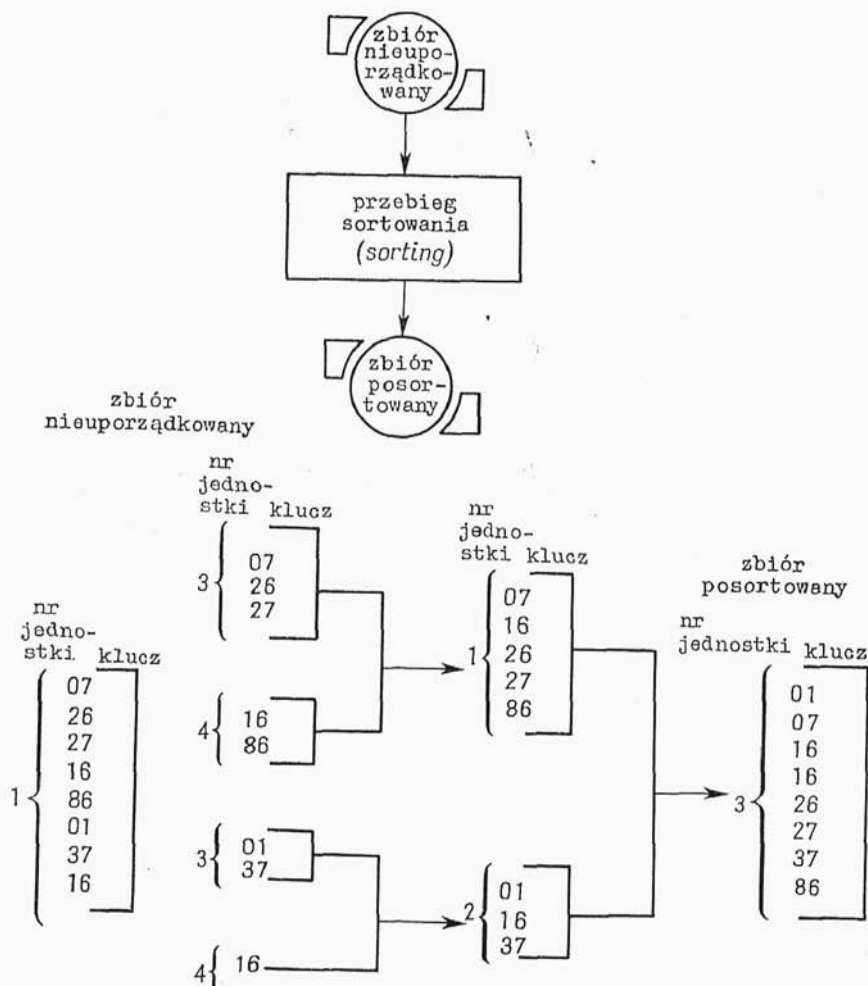
Rys. 62. Przebieg scalania (*merging*) i przykład scalania dwu zbiorów o różnych wartościach kluczy



Rys. 63. Przebieg dobierania (collating) i przykład dobierania rekordów należących do dwu zbiorów na podstawie wartości kluczy

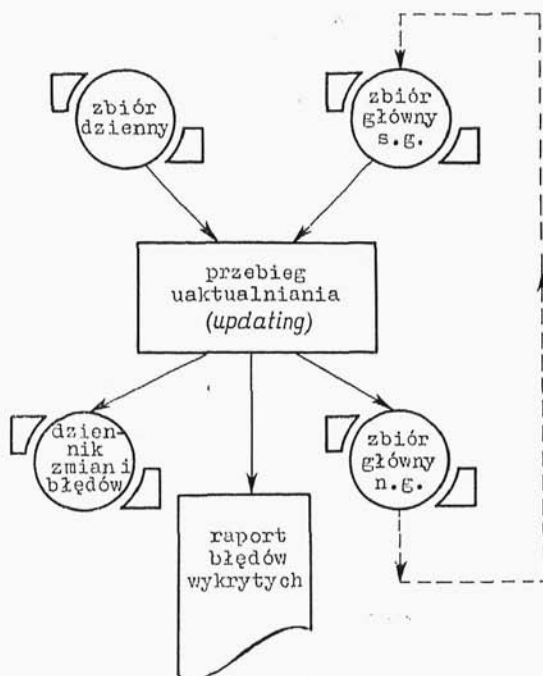
ze względu na wartość kluczy, podzbioru uporządkowanego. Przykład przebiegu dobierania jest pokazany na rys. 63.

Przebieg „sortowania“ (*sorting*) służy do uporządkowania zbioru według wartości kluczy (na ogół w kierunku rosnących wartości



Rys. 64. Przebieg sortowania (*sorting*) i przykład sortowania rekordów zbioru (podano jedynie wartości kluczy) metodą von Neumana przy użyciu czterech jednostek taśmy magnetycznej

kluczy). Przy stosowaniu taśm magnetycznych sortowanie rozpada się na szereg podprzebiegów scalania. Efektywne algorytmy sortowania wymagają posługiwania się co najmniej czterema jednostkami taśmy magnetycznej. Przy stosowaniu czterech jednostek taśmowych zawartość dwu częściowo uporządkowanych taśm (tzw. taśm wejściowych) jest scalana w dłuższe uporządkowane podzbiory umieszczone na przemian na dwu jednostkach wyjściowych. Z chwilą osiągnięcia końca taśm wejściowych, jednostek używanych dotychczas jako wyjściowych zaczynamy używać jako jednostek wejściowych i przeciwnie. Postępowanie takie kontynuujemy aż do uzyskania na jednej taśmie wyjściowej uporządkowanego zbioru (kartoteki). W pierwszym podprzebiegu zbiór nieuporządkowany znajduje się na jednej z taśm wejściowych, druga zaś taśma wejściowa jest pusta. Na rysunku 64 pokazany jest przykład przebiegu sortowania przy użyciu algorytmu von Neumana.

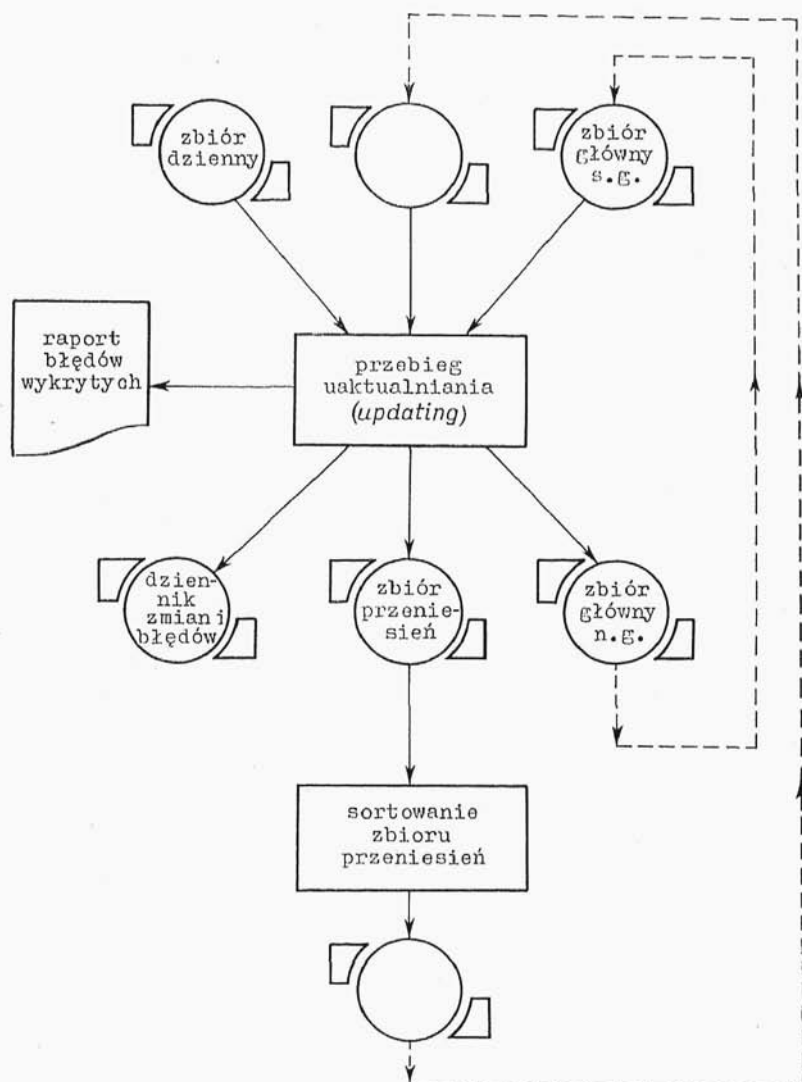


Rys. 65. Przebieg uaktualniania (*updating*) — przypadek uproszczony
s.g. — stara generacja zbioru, n.g. — nowa generacja zbioru.

Przebieg uaktualniania (*updating*). W przebiegu uaktualniania biorą udział dwa zbiory wejściowe zgodnie uporządkowane. Jeden z tych zbiorów, tzw. stara generacja zbioru głównego (*old master file*), jest uzupełniona informacjami z drugiego zbioru wejściowego, tzw. zbioru „dziennego” (*transactions*). W wyniku przebiegu powstaje tzw. nowa generacja zbioru głównego (*new master file*), zbiór pomocniczy sald częściowych przed i po uaktualnieniu zbioru głównego i w tym przebiegu dopiero wykrytych błędów formalnych, tzw. dziennik zmian i błędów (*change journal*) oraz drukowany jest raport wykrytych błędów. Na rysunku 65 pokazany jest schematycznie przebieg uaktualniania.

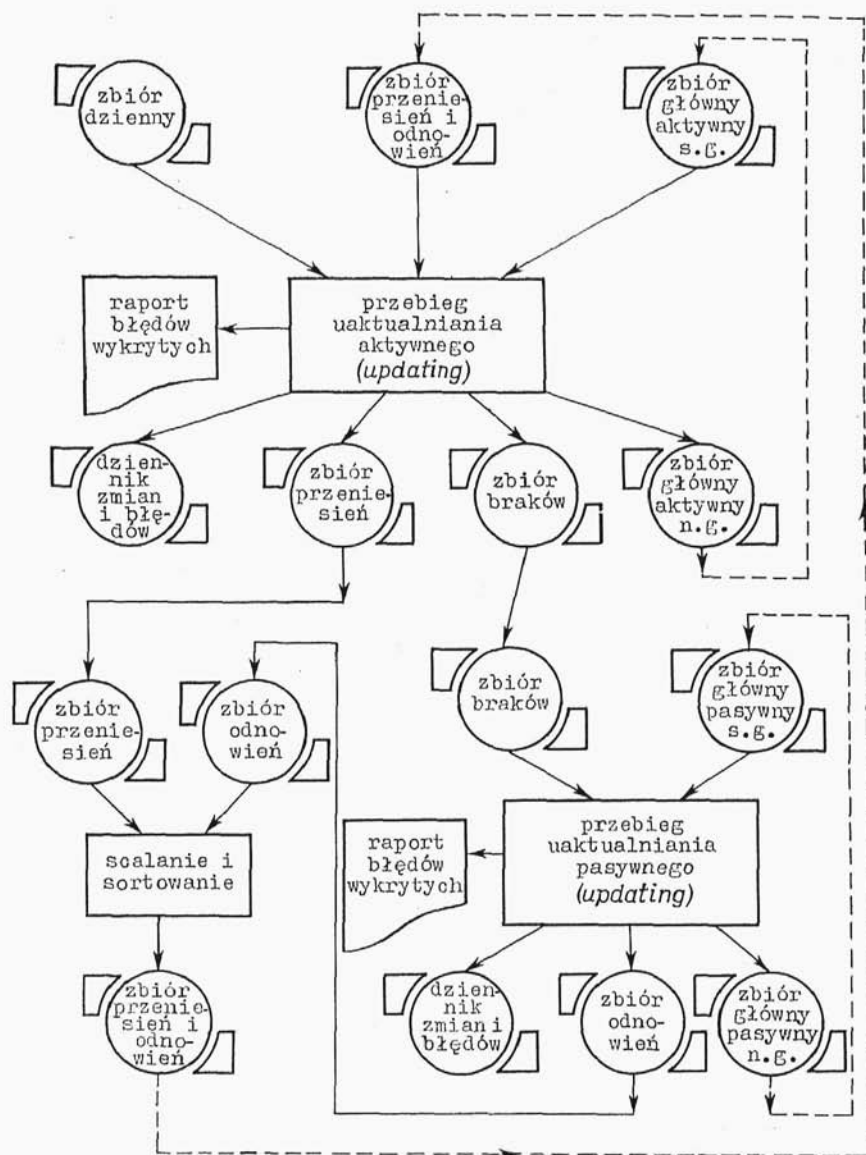
Przebieg uaktualniania z przeniesieniami (*updating with transfers*). Omawiając przebieg sortowania przemilczeliśmy założenie, że zmiany dokonywane w czasie uaktualniania dotyczą poszczególnych rekordów i nie pociągają za sobą konieczności dokonywania zmian w rekordach powiązanych pewnymi zależnościami z danym rekordem. Jeśli tego rodzaju zmiany muszą być dokonywane w rekordach, które w danym uporządkowaniu poprzedzają rekord bezpośrednio uaktualniony za pomocą zbioru dziennego, to zachodzi konieczność gromadzenia w osobny zbiór tych przeniesień, które po posortowaniu są podstawą uaktualniania świeżo powstałej generacji zbioru głównego. W poszczególnych przypadkach tego rodzaju proces może być powtarzany kilkakrotnie, aż do chwili, kiedy nie wystąpi ani jedno przeniesienie. Na rysunku 66 pokazany jest schematycznie przebieg uaktualniania z przeniesieniami. Oczywiście, przy projektowaniu systemu przetwarzania danych, konieczne jest ograniczanie do minimum przebiegów uaktualniania z przeniesieniem. Warto jeszcze w tym miejscu podkreślić, że w przypadku występowania w systemie dużej ilości tego rodzaju przebiegów, pożądane jest zamiast jednostek taśm magnetycznych stosować urządzenia pamięciowe o bezpośrednim dostępie (np. dyski magnetyczne).

Przebieg uaktualniania aktywno-pasywny (*active-inactive updating*). W przypadku gdy w zbiorze głównym występują duże ilości rekordów, co do których można z dużym prawdopodobieństwem przewidzieć, iż w okresie najbliższych kilkunastu lub więcej przebiegów uaktualniania nie będą uaktualniane, można podzielić zbiór główny na dwie części. Na jedną z tych części będą się składały wszystkie te rekordy, które najprawdopodobniej będą uaktualniane w najbliższych przebiegach,



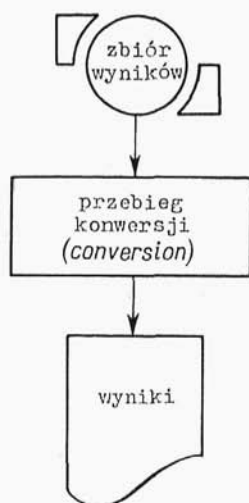
Rys. 66. Przebieg uaktualniania z przeniesieniami (*updating with transfers*)

s.g. -- stara generacja zbioru, n.g. — nowa generacja zbioru.



Rys. 67. Przebieg uaktualniania aktywno-pasywny (active-inactive updating)
 s.g. — stara generacja zbioru, n.g. — nowa generacja zbioru.

i tę część będziemy nazywali zbiorem głównym aktywnym. Na drugą część złożą się pozostałe rekordy i tę część będziemy nazywali zbiorem głównym pasywnym. W trakcie uaktualniania zbioru głównego aktywnego zachodzi konieczność gromadzenia tych rekordów ze zbioru dziennego, dla których nie występują korespondujące rekordy w zbiorze głównym aktywnym i ewentualnie te rekordy zbioru głównego aktywnego, co do których na podstawie informacji ze zbioru dziennego lub innych kryteriów podjęto decyzję przeniesienia ich do zbioru głównego pasywnego. Ten pomocniczy zbiór nazywamy w skrócie zbiorem braków. Na rysunku 67 pokazano schematycznie przebieg uaktualniania aktywno-pasywnego.



Przebieg „konwersji: taśma magnetyczna—urządzenia wyjściowe EMC” (*conversion*). Przebieg ten ma na celu wydrukowanie lub perforację na podstawie wcześniej przygotowanych wyników odpowiednich dokumentów. Na rysunku 68 pokazany jest schematycznie przebieg konwersji: taśma magnetyczna—drukarka liniowa.

Na tym kończymy omawianie najważniejszych przebiegów, składających się na cykle przetwarzania sekwencyjnego.

Rys. 68. Przebieg konwersji taśma magnetyczna—drukarka liniowa (*conversion*)

D.3. PRZETWARZANIE INDYWIDUALNE

Ograniczymy się do bardzo zwięzłego omówienia przetwarzania indywidualnego. Jak już wspominaliśmy, przetwarzanie indywidualne oparte jest na urządzeniach pamięciowych o bezpośrednim dostępie (*random access*), a nie na urządzeniach sekwencyjnych, takich jak taśma magnetyczna. Typowym przykładem urządzeń o bezpośrednim dostępie są dyski magnetyczne. W odróżnieniu od taśm magnetycznych, dyski magnetyczne umożliwiają