

### 3

Przedmiotem działania programów przetwarzania są dane. Te spośród danych, które zostaną wyselekcjonowane w wyniku działania programu (z ewentualnym ich dalszym przetworzeniem), będziemy nazywali *informacjami*. W procesie przetwarzania występuje kilka podstawowych typów danych. Postaramy się możliwie dokładnie omówić każdy z nich. Poza podziałem danych na typy, istotne jest sklasyfikowanie struktur złożonych wyrażeń składających się z danych elementarnych.

Rozważania nasze zaczniemy od podstawowej uwagi. Mianowicie mamy zwykle do czynienia z dwoma głównymi klasyfikacjami struktur [2]. Jedna z nich jest konsekwencją właściwości sprzętu do przetwarzania danych, tzw. *klasyfikacja komputerowo zorientowana*. Druga wynika z potrzeby opisu danych dla całokształtu procesu przetwarzania, tzw. *klasyfikacja procesowo zorientowana*.

W praktyce zwykle niezbyt precyzyjnie formułujemy nasze myśli, co prowadzi do mieszanina pojęć zaczerpniętych z tych dwu klasyfikacji.

W klasyfikacji komputerowo zorientowanej rozróżniamy następujące jednostki danych:

- bit (czyli cyfra liczby przedstawionej w systemie dwójkowym, skrót od pary słów w języku angielskim *binary digit*),

- znak lub byte (znak składa się z sześciu bitów, zaś byte — z ośmiu bitów),

- słowo (w różnych maszynach spotykamy słowa o różnej długości, istnieją również komputery o słowach zmiennej długości),

- blok fizyczny na nośniku wejściowym, wyjściowym lub pamięci zewnętrznej,

- szpula taśmy magnetycznej (pamięć zewnętrzna na taśmie magnetycznej) lub pakiet dyskowy (pamięć zewnętrzna na dyskach wymiennych).

W klasyfikacji procesowo zorientowanej wyróżniamy następujące jednostki danych<sup>1</sup>:

- znak (numeryczny, alfabetyczny lub alfanumeryczny),

- pole,

- grupa pól,

- rekordy (informacyjne, wieloblokowe i identyfikacyjne),

- zbiór,

- bank danych systemu.

### 3.1. ZNAK

*Znak* jest najmniejszą jednostką danych przetwarzania. Znak może być numeryczny dziesiętny, alfabetyczny, alfanumeryczny i binarny. Znak numeryczny

---

<sup>1</sup> Uporządkowane od najmniejszych jednostek do największych.

może przyjmować jedną z dziesięciu wartości cyfrowych: 0, 1, 2, 3, 4, 5, 6, 7, 8, 9; znak alfabetyczny — jedną z „wartości” alfabetu od A do Z; znak alfanumeryczny — jedną z „wartości” spośród dziesięciu cyfr dziesiętnych, liter alfabetu, albo symboli specjalnych, jak np.: +, -, ·, =, :, /, przecinek, % itd.; znak binarny (dwójkowy) — ma „wartość” numeryczną swojego kodu dwójkowego. O nadaniu takiego czy innego znaczenia kodowi dwójkowemu, przechowywanemu w pamięci komputera, decyduje program operujący danym kodem. Może się zdarzyć, że w ramach danego procesu przetwarzania danemu kodowi dwójkowemu są nadawane różne znaczenia.

Czytelnika może zdziwić używanie terminu *wartość* w odniesieniu do liter i symboli specjalnych. Termin ten został użyty celowo, ze względu na przyporządkowanie wewnątrz komputera każdej cyfrze dziesiętnej, każdej literze alfabetu i każdemu symbolowi specjalnemu jednoznacznie określonego kodu dwójkowego, o jednoznacznie określonej liczbie bitów (np. sześciu lub ośmiu).

### 3.2. POLE

*Pole* jest to skończony ciąg znaków przeznaczony do zapisania informacji kodowej za ich pomocą.

Przez *pole numeryczne* będziemy rozumieli ciąg znaków numerycznych. Pojęcie pola numerycznego identyfikujemy z pojęciem liczby dziesiętnej o określonej liczbie cyfr dziesiętnych.

Przez *pole alfabetyczne* będziemy rozumieli ciąg znaków alfabetycznych. Jako przykład może tu posłużyć pole przeznaczone do zapisywania nazwiska.

Przez *pole alfanumeryczne* będziemy rozumieli ciąg znaków alfanumerycznych. Przykładem może tu być pole przeznaczone do zapisywania adresu zamieszkania. Oczywiście jest, że każde pole alfabetyczne jest równocześnie polem alfanumerycznym. Natomiast twierdzenie odwrotne nie jest prawdziwe.

Przez *pole binarne* będziemy rozumieli ciąg znaków binarnych. Pojęcie pola binarnego identyfikujemy z pojęciem liczby dwójkowej o określonej liczbie bitów.

*Rozmiar pola* jest to maksymalna ilość znaków, jaka może być zapisana w danym polu. Przyjmuje się zwykle konwencję, że pola numeryczne są adiustowane do prawego brzegu pola (czyli do najmniej znaczącej pozycji cyfrowej), natomiast pola alfabetyczne i alfanumeryczne — do lewego brzegu pola.

### 3.3. GRUPA PÓL

*Grupa pól* jest to zespół kilku pól tworzących całość z punktu widzenia danego procesu przetwarzania. Na przykład pola: nazwisko i imię oraz pole: adres zamieszkania tworzą grupę pól z punktu widzenia identyfikacji występującej w procesie przetwarzania, omawianym przykładowo w rozdziale 1.

### 3.4. REKORD INFORMACYJNY

*Rekord informacyjny*, zwany często *rekordem*, jest odpowiednikiem dokumentu jednostkowego lub pozycji dokumentu wielopozycyjowego. Rekord może być przechowywany na różnych nośnikach maszynowych. Mamy

zwykle do czynienia z rekordami kartowymi, zwanymi krótko kartami, rekordami przechowywanymi na taśmach magnetycznych, na dyskach wymiennych, w pamięci operacyjnej komputera i wreszcie na drukarce liniowej, czyli z liniami tabulogramu, z których każda linia jest traktowana jako oddzielny rekord. Rekord zawiera przynajmniej jedno pole. Z punktu widzenia procesu przetwarzania rozróżniamy trzy rodzaje rekordów informacyjnych: 1) wejściowe, 2) wyjściowe i 3) wejściowo-wyjściowe.

*Rekordy informacyjne wejściowe* są to rekordy wprowadzane do pamięci operacyjnej komputera z urządzeń wejścia, np. z czytnika kart perforowanych; w tym przypadku są to równocześnie rekordy kartowe. Rekordy wejściowe składają się z reguły z dwu podstawowych grup pól:

1) grupy pól przeznaczonej do identyfikowania rekordu, tzw. cechy identyfikacyjnej, często zwanej *kluczem rekordu*,

2) grupy pól informacyjnych nie wchodzących do klucza, często zwanej *ciałem rekordu*.

*Rekordy informacyjne wyjściowe* są to rekordy wprowadzane z pamięci operacyjnej komputera do urządzenia wyjścia, np. na drukarkę liniową; w tym przypadku są to linie tabulogramu. Rekordy wyjściowe często nie mają klucza, składają się więc z zasadzie z jednej grupy pól.

*Rekordy informacyjne wejściowo-wyjściowe* są to rekordy wprowadzane lub wyprowadzane do i z pamięci operacyjnej komputera z i do jednostek pamięci zewnętrznej, w kolejnych fazach przebiegu. Rekordy te, podobnie jak rekordy informacyjne wejściowe, składają się z reguły z dwu pól (klucza i ciała).

Dotychczas nie omawialiśmy sposobu przechowywania rekordów na nośnikach maszynowych zewnętrznych, takich jak: karta perforowana i taśma magnetyczna. Zgodnie z wprowadzoną komputerowo zorientowaną klasyfikacją powinniśmy zająć się bliżej relacją pomiędzy rekordem informacyjnym a blokiem fizycznym na danym typie nośnika zewnętrznego.

Jeśli nośnikiem jest karta perforowana, blok fizyczny ma rozmiar równy jednej karcie (np. o zawartości maksymalnej 80 znaków alfanumerycznych, czyli tzw. obraz karty 80-kolumnowej). Przyjmujemy wówczas konwencję następującą: jedna karta perforowana (jeden blok fizyczny) służy do przechowywania jednego rekordu informacyjnego.

W przypadku taśmy magnetycznej sprawa jest bardziej złożona. Rozmiary bloków fizycznych na taśmie magnetycznej, ze względu na optymalne wykorzystanie nośnika, powinny mieć rozmiar rzędu kilkuset lub nawet kilku tysięcy znaków. Z drugiej strony, jednostkowe rekordy informacyjne (np. odpowiedniki kart perforowanych) mają rozmiary przeważnie rzędu setki znaków. Dlatego też dla przechowywania zbiorów sekwencyjnych, złożonych z jednostkowych rekordów informacyjnych, przyjmujemy zwykle następującą konwencję: jeden blok fizyczny służy do przechowywania wielu rekordów informacyjnych. Dla zwiększenia efektywności wykorzystania nośnika wprowadzamy rekordy informacyjne o zmiennej długości. Każdy rekord musi posiadać pole, w którym podana jest jego długość, liczona w znakach lub słowach (w zależności od specyfiki organizacyjnej danego typu komputera). Pole zawierające długość danego rekordu musi we wszystkich rekordach wszystkich zbiorów sekwencyj-

nych taśmowych znajdować się zawsze w tym samym miejscu. Jest to niezbędne przy budowie zunifikowanego pakietu programowego — współpracy z taśmami magnetycznymi (*magnetic tape house keeping package*).

### 3.5. REKORDY WIELOBLOKOWE I PSEUDOREKORDY

Do tej pory ograniczaliśmy się do rekordów informacyjnych jednostkowych, będących odpowiednikiem dokumentów jednostkowych. Jednakże w bardziej złożonych zastosowaniach (np. kierowanie procesem wytwórczym) zachodzi konieczność operowania wewnątrz systemu rekordami (a właściwie podzbiorami o bezpośrednim dostępie) zawierającymi w sobie wiele tablic o rozmiarach przekraczających wielkość bloków fizycznych na taśmach magnetycznych. Tego rodzaju rekordy są odpowiednikiem wielopozycyjnych dokumentów. W maszynach o bardzo dużej pamięci operacyjnej posługiwanie się dużymi rekordami wielopozycyjnymi jest stosunkowo łatwe. Natomiast w maszynach o małej pamięci operacyjnej, wyposażonych dodatkowo w pamięć bębnową lub inną pamięć zewnętrzną o bezpośrednim dostępie, zachodzi konieczność przenoszenia rekordu do pamięci bębnowej i traktowania go jako podzbioru o dostępie bezpośrednim.

Rekord wielopozycyjny może przekraczać rozmiarami pamięć operacyjną. W tym przypadku nawet blok fizyczny o maksymalnym rozmiarze nie wystarcza do przechowywania jednego rekordu wielopozycyjnego. Mówimy wówczas, że mamy do czynienia z tzw. *rekordami wieloblokowymi*. Rekordy takie dzielimy na

pseudorekordy, o rozmiarach równych przeciętnemu rekordowi jednostkowemu, i przechowujemy na taśmie magnetycznej w formie sekwencji pseudorekordów. Rekord wielopozycyjny przepisujemy najpierw do pamięci o dostępie bezpośrednim lub do obszaru pamięci operacyjnej (pseudorekord po pseudorekordzie) i dopiero wtedy traktujemy go jak podzbiór o dostępie bezpośrednim. Podobnie rzecz się przedstawia, jeśli taki rekord chcemy zapisać na taśmie magnetycznej: najpierw dzielimy go na pseudorekordy, a następnie zapisujemy pseudorekordy jeden po drugim na taśmie magnetycznej. Zbiór złożony z rekordów wieloblokowych będziemy w dalszym ciągu nazywali *zbiorem złożonym z sekwencji podzbiorów o dostępie bezpośrednim*.

### 3.6. REKORD IDENTYFIKACYJNY

Oprócz rekordów informacyjnych wyróżniamy rekordy pomocnicze — identyfikacyjne, zwane *etykietami*. Przyjmuje się następującą konwencję identyfikowania taśm magnetycznych: na początku taśmy magnetycznej — rekord identyfikacyjny (tzw. etykieta początku szpuli), następnie — kolejne rekordy informacyjne, zaś na końcu taśmy magnetycznej — drugi rekord identyfikacyjny (tzw. etykieta końca szpuli).

Celem etykiet jest umożliwienie identyfikowania, przez specjalny program organizacyjny komputera, zawartości danego nośnika pamięci zewnętrznej. Jest to, między innymi, podyktowane koniecznością zabezpieczenia przed wykorzystaniem niewłaściwych, w ramach danego procesu informacyjnego, wymiennych nośników pamięci zewnętrznych.



### 3.7. ZBIÓR DANYCH

Wprowadzenie pojęcia zbioru wymaga wielu dodatkowych wyjaśnień. Dlatego też zastosujemy tu jak gdyby metodę kolejnych przybliżeń. W sposób bardzo ogólny możemy określić zbiór danych jako zespół formatów rekordów i kryteriów ich uporządkowania. Definicja taka, mimo że formalnie poprawna, jest trudna do zrozumienia. W praktyce mamy do czynienia nie ze zbiorami, ale z kolejnymi generacjami zbiorów. Zbiór jest pojęciem dynamicznym, dla którego określono jedynie ramy. W toku operacji, zwanej aktualizacją zbioru (patrz s. 119), kolejną generację zbioru zastąpimy przez następną, w danym momencie aktualniejszą. Uwaga ta — jak się dalej przekonamy — dotyczy jedynie tzw. zbiorów wewnętrznych systemów, aktualizowanych automatycznie w ramach danego systemu elektronicznego przetwarzania danych. Dla większego uporządkowania naszych rozważań wprowadzimy następującą klasyfikację zbiorów danych dla systemów przetwarzania:

— zbiory wejściowe zewnętrzne, dla których typowymi nośnikami są karty perforowane, taśma papierowa perforowana itp.;

— zbiory wyjściowe zewnętrzne, dla których typowym nośnikiem jest papier w drukarce wierszowej (w tym przypadku tabulogramy);

— zbiory podstawowe systemu, dzielące się dalej, ze względu na przebiegi przetwarzania danego systemu, na: zbiory podstawowe aktualizowane (wejściowo-wyjściowe), zbiory podstawowe wejściowe (bierne), zbiory podstawowe generowane (wyjściowe); zgodnie z do-

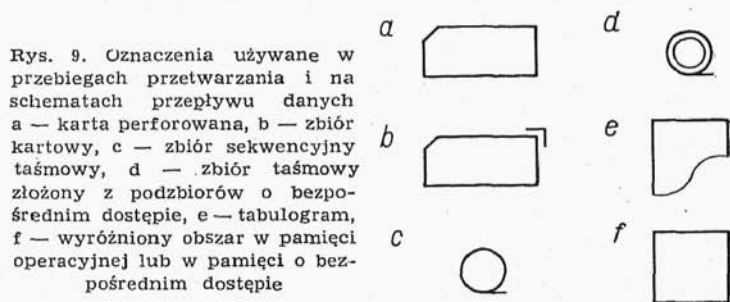
tychczasowymi rozważaniami, zbiory podstawowe przechowywane na taśmach magnetycznych podzielimy na: zbiory złożone z indywidualnych rekordów (zbiory sekwencyjne) i zbiory złożone z podzbiorów o bezpośrednim dostępie;

— zbiory robocze, z których każdy, z punktu widzenia jednego przebiegu przetwarzania, jest zbiorem wyjściowym wewnętrznym systemu i równocześnie, z punktu widzenia co najmniej jednego przebiegu przetwarzania, jest zbiorem wejściowym wewnętrznym systemu.

Zbiory podstawowe i zbiory robocze wewnętrzne obejmujemy wspólną nazwą zbiorów wewnętrznych systemu. Podstawowymi nośnikami dla zbiorów wewnętrznych są nośniki urządzeń pamięci zewnętrznych komputera.

W stosunku do zbiorów zewnętrznych i zbiorów roboczych używamy pojęcia generacji zbioru. Jednakże sposób powstawania kolejnych generacji tych zbiorów wygląda inaczej niż w przypadku zbiorów podstawowych.

Na rysunku 9 przedstawiamy podstawowe oznaczenia elementów opisu danych.



Przez *generację zbioru* rozumiemy dany zestaw rekordów, z których każdy ma format dopuszczalny w ramach danego zbioru danych. Przyjmujemy przy tym, że kolejne generacje zbioru powstają w zależności od jego kategorii. I tak:

- kolejna generacja zbioru wejściowego zewnętrznego powstaje na podstawie partii dokumentów źródłowych dotyczących ściśle określonego przedziału czasu,

- kolejna generacja zbioru wyjściowego zewnętrznego powstaje w wyniku działania pewnej części systemu (tzw. cyklu przetwarzania) na podstawie danej generacji zbiorów wewnętrznych i ewentualnie zbiorów wejściowych zewnętrznych,

- kolejna generacja zbioru podstawowego powstaje w wyniku aktualizacji poprzedniej generacji danego zbioru podstawowego określonymi generacjami zbioru roboczego lub zbioru wejściowego zewnętrznego,

- kolejna generacja zbioru roboczego wewnętrznego powstaje w wyniku działania ustalonej części systemu (tzw. przebiegu przetwarzania) w oparciu o dane generacje zbiorów wewnętrznych i ewentualnie zbiorów wejściowych zewnętrznych.

### 3.8. BANK DANYCH SYSTEMU

Przez *bank danych* rozumiemy zespół wszystkich aktualnych generacji zbiorów podstawowych systemu. W ramach niniejszego rozdziału nie omawiamy tego pojęcia szczegółowo, ponieważ poświęcamy mu oddzielny rozdział.