

maszyny matematyczne

zastosowania w gospodarce, technice i nauce

Nr 9

MIESIĘCZNIK

1 9 6 8

R O K I V

W r z e s i e ń

Organ Pełnomocnika Rządu do Spraw Elektronicznej Techniki Obliczeniowej i Naczelnej Organizacji Technicznej



TADEUSZ WALCZAK

Główny Urząd Statystyczny
Warszawa

Dr Tadeusz Walczak, dyrektor Ośrodka Elektronicznego Głównego Urzędu Statystycznego, doktor nauk ekonomicznych. Obronił pracę doktorską w roku 1962 z zakresu mechanicznego przetwarzania danych. W dziedzinie tej pracuje wiele lat. Jest autorem kilku książek i szeregu publikacji w czasopiśmie na temat maszyn liczących i organizacji zmechanizowanego przetwarzania danych.

681.322.004.14:31

Zastosowanie EMC do przetwarzania danych statystycznych

W artukule opisano zastosowanie maszyny ICT-1905 do prac statystycznych w Głównym Urzędzie Statystycznym PRL. Podano zestaw maszyny, ogólny schemat przetwarzania i tematykę prac. Przedstawiono organizację procesu przetwarzania danych na EMC, ze szczególnym uwzględnieniem programów kontroli i aktualizacji danych. Autor podkreśla trudności pracy ośrodka obliczeniowego, związane z masowością informacji oraz z koniecznością częstych zmian systemów i programów. Przeprowadza analizę wykorzystania czasu pracy EMC i przekazuje wnioski z wdrażania pracy wieloprogramowej.

Doskonalenie procesu planowania i zarządzania gospodarką narodową uzależnione jest w dużym stopniu od możliwości otrzymywania dokładnej i terminowej informacji o najważniejszych zjawiskach zachodzących w gospodarce.

Najważniejszą rolę do spełnienia ma w tej dziedzinie informacja statystyczna, badająca stan i rozwój zjawisk zachodzących w procesie produkcji i wymiany, charakteryzująca pod względem ilościowym wszystkie ważniejsze przejawy działalności produkcyjnej i kulturalnej społeczeństwa oraz procesy demograficzne w społeczeństwie.

Z uwagi na to, że badania statystyczne przeważnie opierają się na masowym materiale jednostkowym, opracowanie materiałów badań statystycznych jest procesem niezwykle pracochłonnym.

Z tego względu unowocześnienie techniki przetwarzania danych statystycznych jest przedmiotem szczególnego zainteresowania we wszystkich krajach.

Nie jest dziełem przypadku, że właśnie potrzeby statystyki doprowadziły do skonstruowania przez H. Holleritha w roku 1888 — pierwszych maszyn licząco-analitycznych na zasadzie automatycznego odczytu danych z kart dziurkowanych podobnie, jak nie przypadkiem pierwsza elektroniczna maszyna cyfrowa do przetwarzania danych została zastosowana właśnie do prac statystycznych¹⁾.

W Polsce technika przetwarzania danych statystycznych przez długi okres pozostawała w tyle, zarówno w stosunku do urzędów statystycznych za granicą, jak i w stosunku do stale narastających potrzeb w zakresie opracowania informacji statystycznej, odpowiadającej wymaganiom planowania i zarządzania.

Dopiero w ostatnich latach nastąpił dość szybki rozwój w tej dziedzinie, wyrażający się w zwiększeniu dostaw maszyn liczących oraz organizacji kilkunastu

¹⁾ W marcu 1951 roku zainstalowano EMC UNIVAC I w amerykańskim Urzędzie Statystycznym.

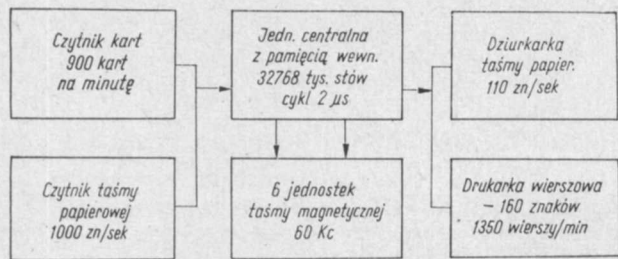
ośrodków maszyn w poszczególnych ośrodkach wojewódzkich.

Technika elektroniczna w GUS

Szczególnie ważny krok na drodze unowocześnienia techniki przetwarzania danych statystycznych stanowi zainstalowanie w Głównym Urzędzie Statystycznym w październiku 1967 roku — elektronicznej maszyny cyfrowej ICT-1905. Maszyna ta jest największa i najszybsza spośród maszyn zainstalowanych dotąd w naszym kraju.

Zastosowanie maszyny elektronicznej w pracach statystycznych umożliwi nie tylko przyspieszenie opracowania wyników badań, lecz pozwoli również rozwiązać w sposób nowoczesny szereg najistotniejszych problemów organizacji, zbierania i obiegu informacji ekonomicznej w kraju.

Maszyna zainstalowana w GUS pracuje w zestawie przedstawionym w tablicy.



Jednostkowe dane statystyczne podlegające opracowaniu wprowadza się do maszyny za pomocą kart dziurkowanych z teoretyczną prędkością 900 kart na minutę lub za pomocą taśmy dziurkowanej z prędkością 1000 znaków na sekundę.

Wprowadzane do maszyny dane opracowywane są przez jednostkę centralną z bardzo dużą prędkością. Przykładowo — operacje dodawania i odejmowania maszyna wykonuje z prędkością ponad 142 tys., mnożenia 25 tys., dzielenia 23 tys. operacji na sekundę. Pozwala to wykonywać bardziej skomplikowane obliczenia statystyczne oraz prowadzić bardziej szczegółową kontrolę logiczną danych jednostkowych.

Wprowadzone do maszyny dane jednostkowe zapisuje się na taśmy magnetyczne. Przechowywanie danych na taśmach magnetycznych jest bardzo wygodne, zarówno ze względu na dużą prędkość odczytu wynoszącą 60 tys. znaków na sekundę, jak i na dużą gęstość zapisu pozwalającą przechowywać ogromne ilości danych na niewielkich pod względem objętości krążkach taśmy.

Na jednym krążku taśmy magnetycznej o średnicy 30 cm i ważącym ok. 1,8 kg można zapisać informację zawierającą ok. 10 mln znaków, co odpowiada minimum 125 tys. kart dziurkowanych, których ciężar wynosi ponad 370 kg.

Wyniki obliczeń wykonywanych przez maszynę wprowadza się w formie zestawień tabelarycznych za pomocą drukarki wierszowej w wałek drukujący o 160 znakach, pozwalający sporządzać dowolne zestawienia o maksymalnej szerokości 41 cm. Szybkość pracy drukarki wynosi 1100—1350 wierszy na minutę.

Tematyka opracowań wykonywanych na EMC

Program pracy Ośrodka Elektronicznego obejmuje szereg najbardziej pracochłonnych tematów opracowań, zarówno z dziedziny tzw. statystyki bieżącej, opartej na sprawozdawczości i przewidzianej w rocznym programie prac statystycznych, jak i przetwarzaniu materiałów jednorazowych badań masowych, prowadzonych przez organy statystyki państwowej.

Opracowywana przez EMC informacja dotyczy między innymi: obrotów handlu zagranicznego, migracji lud-

ności (zmiany miejsca zamieszkania), zgonów, działalności przedsiębiorstw przemysłowych, efektywności postępu technicznego w przemyśle, zatrudnienia ludność, budżetów rodzinnych itp. Przygotowywane są także programy opracowania materiałów spisu pracowników ze średnim i wyższym wykształceniem oraz dojazdów do miejsca pracy z miejsca zamieszkania. Prócz tego wykonuje się także szereg obliczeń matematyczno-statystycznych związanych ze stosowaniem metod statystyki matematycznej w opracowaniach Głównego Urzędu Statystycznego.

Organizacja opracowań statystycznych na maszynie elektronicznej

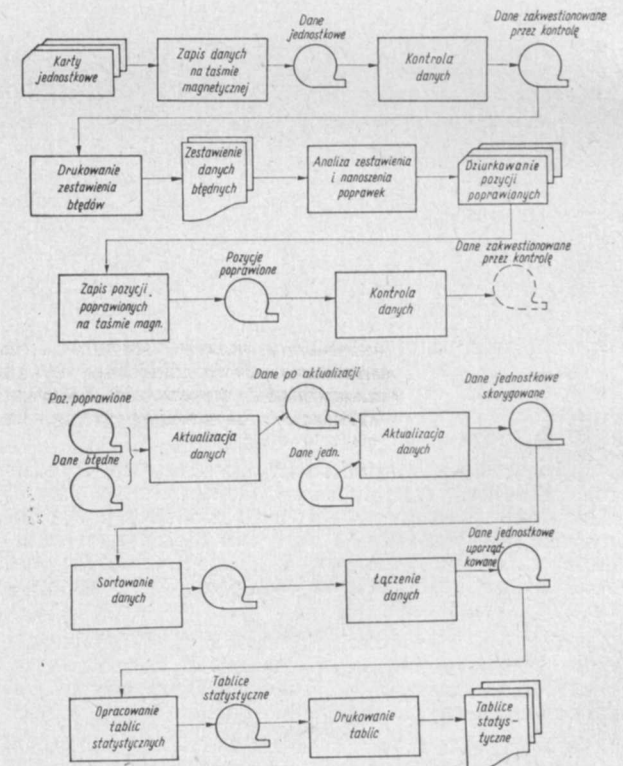
Przetwarzanie informacji statystycznej na maszynie elektronicznej odznacza się szeregiem właściwości szczególnych, różniących te opracowania od przetwarzania danych w innych dziedzinach. Właściwości te wywierają wpływ na sposób opracowania programów, strukturę wykorzystania czasu maszyny, gospodarkę taśmami magnetycznymi itp.

Przebieg przetwarzania danych statystycznych na maszynie elektronicznej w najogólniejszym zarysie przedstawia rysunek.

Informację źródłową podlegającą opracowaniu przynosi się z dokumentów źródłowych (sprawozdań, ankiet itp.) na karty dziurkowane. Następnie karty dziurkowane wprowadza się do maszyny elektronicznej w celu przeniesienia informacji wydziurkowanej na kartach na taśmę magnetyczną. Równocześnie z zapisem danych na taśmę dokonuje się sprawdzenia liczby kart w celu ustalenia kompletności zbiorów.

Przeniesione na taśmę magnetyczną dane jednostkowe mogą zawierać — (i niestety prawie zawsze zawierają) szereg błędów. Błędy te wynikają z niewłaściwego wypełniania dokumentów przez sprawozdawców, błędnego kodowania lub błędnego wydziurkowania kart. Z tego względu informacja ta przed rozpoczęciem opracowania wyników poddawana jest szczegółowej kontroli.

Konieczność przeprowadzenia szczegółowej kontroli danych jednostkowych stanowi jedną z cech szczególnych przetwarzania informacji statystycznej.



Ogólny schemat przetwarzania danych statystycznych

Głównym zadaniem tej kontroli jest wyeliminowanie z materiału źródłowego rażących błędów, mogących zniekształcić wyniki badań statystycznych lub takich danych błędnych, które uniemożliwiają sporządzanie tablic lub też powodują pojawienie się w tablicach wynikowych danych absurdalnych, mogących podważać zaufanie czytelnika do publikacji statystycznych.

Programy automatycznej kontroli danych jednostkowych są zwykle bardzo obszerne i skomplikowane. Przykładowo można podać że program automatycznej kontroli informacji jednostkowej z zakresu budżetów rodzin pracowniczych zawiera ok. 5000 rozkazów i zajmuje 25 000 słów w pamięci wewnętrznej. W większości tematów opracowań program kontroli przewiduje sprawdzenie danych jednostkowych z różnych punktów widzenia, a w szczególności:

- 1) z punktu widzenia formalnego
- 2) z punktu widzenia logicznego
- 3) z punktu widzenia zgodności arytmetycznej.

Za błąd formalny uważa się przykładowo wystąpienie znaku literowego w miejscu, gdzie powinna być wyłącznie informacja cyfrowa, wystąpienie nie istniejącego symbolu, brak zapisu w miejscu, gdzie musi on występować itp.

Kontrola logiczna ma na celu wykrycie w materiale źródłowym zapisów, poszczególne elementy, które nie mieszczą się w ustalonych z góry granicach lub też znajdują się w logicznej sprzeczności w stosunku do siebie. Przykładowo w opracowaniu materiałów spisu kadr z wyższym i średnim wykształceniem program kontroli logicznej sygnalizuje błąd (przypadek wątpliwy), jeśli wśród zatrudnionych wystąpi pracownik młodszy od 16 lub starszy od 80 lat; jeśli staż pracy jest wyższy niż 40 lat, jeśli pracownik z wyższym wykształceniem ma mniej niż 21 lat itp.

Podobnie jako błędny zostanie uznany zapis, jeśli dla pracownika posiadającego symbol wyższego wykształcenia w miejscu przeznaczonym na zapis zajmowanego stanowiska wystąpi zapis „pracownik fizyczny”, jeśli pracownik posiada tytuł naukowy, a w miejscu przeznaczonym na zapis wykształcenia wystąpi symbol wykształcenia inny niż „wyższe” itp.

Program automatycznej kontroli przewiduje zapis zakwestionowanych przez maszynę danych na odrębnej taśmie magnetycznej, na podstawie której następnie drukuje się zestawienie pozycji błędnych.

W zestawieniu tym drukuje się pełną treść danej pozycji jednostkowej, odpowiadającej ściśle treści wydziurkowanej karty oraz przy każdej pozycji odpowiednie symbole błędów. Przykładowo, jeśli w danej pozycji jednostkowej maszyna stwierdzi nie istniejący symbol przedsiębiorstwa, to oprócz pełnej treści danej pozycji pisze symbol „01”, jeśli wystąpi nielogiczne powiązanie pomiędzy wiekiem i zawodem — symbol „02” itp.

Wykaz pozycji zakwestionowanych przez maszynę jest następnie poddawany szczegółowej analizie przez statystyków, poszczególne pozycje wykazu konfrontowane są z zapisem dokumentów źródłowych i jednocześnie wprowadza się w poszczególnych pozycjach wykazu niezbędne poprawki. Jeśli w wykazie znajdują się pozycje, które maszyna uznała jako wątpliwe lub nietypowe, a w wyniku analizy uznane zostaną za poprawne — to takie pozycje zostają skreślone z wykazu. Następnie na podstawie poprawionego wykazu dziurkuje się nowe karty, które wczytuje się i kontroluje podobnie, jak karty pierwotne aż do całkowitego wyeliminowania błędów z materiału jednostkowego.

Następnym etapem opracowania jest tzw. aktualizacja danych. Polega ona na usunięciu ze zbioru pozycji uznanych za błędne i wniesienie na ich miejsce pozycji skorygowanych.

Po skorygowaniu zbioru informacji jednostkowych na taśmie magnetycznej rozpoczyna się właściwe opracowanie danych, a więc sortowanie oraz tworzenie i drukowanie tablic.

W większości przypadków tablice opracowuje się w taki sposób, aby można było przenosić je na klisze

metodą kserograficzną i powielać bez potrzeby przepisywania na maszynie.

Inną ważną cechą, specyficzną dla przetwarzania informacji statystycznej, wywierającą wpływ na organizację opracowań jest bardzo duża masowość informacji jednostkowych dla poszczególnych zbiorów opracowywanych jednorazowo na maszynie elektronicznej.

Przykładowo można podać, że zbiór informacji jednostkowych z zakresu statystyki zgonów składa się z 240 tys. pozycji jednostkowych (rekordów), zbiór z zakresu statystyki budżetów rodzinnych — 600 tys. pozycji, statystyki ruchu wędrownego — ok. 900 tys. pozycji, ankieta warunków mieszkaniowych ludności — ok. 400 tys. pozycji, statystyka kadr ze średnim i wyższym wykształceniem — ok. 3 mln pozycji itd.

Fakt ten oznacza, że każdy zbiór zajmuje praktycznie kilkanaście (lub w niektórych przypadkach nawet kilkadziesiąt) szpul taśm magnetycznych. Zmusza to Ośrodek Elektroniczny do zapewnienia odpowiedniej organizacji biblioteki taśm magnetycznych, dokładnego planowania czasu poszczególnych opracowań, stosowania odpowiednich metod sortowania i łączenia zbiorów wieloszpułowych, zapewnienia możliwości przerywania przebiegu programów z możliwością ich następnego uruchomienia itd.

Duża masowość informacji oraz konieczność jej długotrwałego przechowywania wymaga dużej liczby taśm magnetycznych. Biblioteka taśm magnetycznych zawiera obecnie ok. 600 szpul taśm i mimo to Ośrodek zaczyna odczuwać poważny brak taśm.

Bardzo poważny wpływ na pracę Ośrodka Elektronicznego wywiera fakt, że w opracowaniach statystycznych mamy do czynienia z zasady z pracami jednorazowymi.

Nawet jeśli temat pracy powtarza się z roku na rok, to zmiany zachodzące w gospodarce, w systemie planowania itp. powodują zmiany w zakresie informacji jednostkowej, w układzie zestawień wynikowych, sposobach grupowania danych itd., co pociąga z kolei za sobą konieczność stałego opracowywania nowych programów. Fakt ten zmusza Ośrodek do zatrudnienia wielu projektantów systemu i programistów będących w stanie zaprojektować nowe systemy oraz przygotować dla nich programy w bardzo krótkim czasie.

Z drugiej strony fakt dużej zmienności w programach i konieczności przygotowywania wciąż nowych programów angażuje dużą ilość czasu maszyny na testowanie i uruchamianie programów.

Wykorzystanie maszyny elektronicznej

Maszyna została zainstalowana w Ośrodku Elektronicznym w październiku 1967 roku. W dniu 10 października tego roku, po przejściu prób technicznych, przewidzianych w kontrakcie — maszyna została przyjęta do eksploatacji. W okresie od 11 do 14 października przeprowadzona była nieprzerwana praca maszyny przez 72 godziny, a od 16 października rozpoczęto pracę na dwie zmiany. Obecnie, wobec bardzo dużego nawału prac Ośrodek czyni starania o uruchomienie całodobowej pracy ciągłej. Należy przypuszczać, że do chwili ukazania się niniejszego artykułu w druku — maszyna pracować będzie na trzy zmiany²⁾.

Ogólny czas pracy maszyny za okres od początku roku do 15 maja 1968 r. (dwie zmiany plus kilka dodatkowych zmian nocnych) wyniósł 1580 godzin. Czas ten wykorzystany był w sposób następujący:

● przetwarzanie użytkowe	40,6%
● translacja i uruchomienie programów	38,9%
● etykietowanie nowych taśm, aktualizacja translatorów, modyfikacje techniczne itp. prace własne	2,3%
● konserwacja planowa	14,2%
● przestoje awaryjne	4,0%

2) Już od 10.VI. 1968 r. maszyna pracuje na 3 zmiany.

Czas przetwarzania użytkowego obejmuje czas odczytu danych z kart, sortowanie danych na taśmach magnetycznych, czas opracowania wyników oraz drukowanie tablic wynikowych. Czas ten wynika z podliczeń rejestracji zegara maszyny (maszyna drukuje co minutę na maszynie do pisania zainstalowanej na pulpicie operatorskim dokładny czas). Efektywny czas przetwarzania danych był nieco większy, dzięki wykorzystaniu wieloprogramowości maszyny. Maszyna ICT-1905 może wykonywać równocześnie do 4 programów pod warunkiem, że jest wyposażona w dostateczną liczbę urządzeń peryferyjnych, które można by przydzielić do poszczególnych programów. Przy obecnym, bardzo skromnym wyposażeniu (1 czytnik kart, 1 drukarka wierszowa, 6 jednostek taśm magnetycznych) możliwości pracy wieloprogramowej są bardzo poważnie ograniczone, tym bardziej że opracowywane przez Ośrodek tematy prac są bardzo różnorodne i w większości dość skomplikowane. Mimo to, szczególnie w związku z występującym ostatnio dużym obciążeniem maszyny, wprowadza się, gdzie tylko jest to możliwe, pracę wieloprogramową. Przykładowo — w okresie pierwszej połowy maja br. w ramach wieloprogramowości przepracowano 29 godzin.

W czasie pracy w systemie wieloprogramowym nie stwierdzono żadnych trudności technicznych ani programowych. Można więc stwierdzić, że jedynym ograniczeniem wieloprogramowości jest pojemność pamięci oraz liczba jednostek peryferyjnych i liczba jednostek pamięci taśmowej. Wymaga jest także większa sprawność obsługi operatorskiej oraz wyższy poziom planowania operatywnego.

Dość duży procent czasu maszyny przeznaczają się na tłumaczenie (translacje) programów opracowywanych w języku PLAN, COBOL lub ALGOL na język wewnętrzny maszyny oraz na uruchamianie programów. Jak wspomniano na początku artykułu, jest to zjawisko w głównej mierze obiektywne, wynikające ze specyfiki opracowań statystycznych jako prac w większości niepowtarzalnych i jednocześnie trudnych do zaprogramowania z uwagi na swą różnorodność i wielostronność.

W pierwszym kwartale 1968 r. przetestowano prawie 200 programów, a łączna liczba programoprzebiegów (licząc za 1 przebieg translację oraz realizację pro-

gramu z danymi próbnymi) wyniosła ponad 1100. Tak więc przeciętnie dziennie (w czasie dwóch zmian) testowano 15 programów, a przeciętny czas testowania 1 programu wyniósł 15—20 minut.

Czas konserwacji utrzymujący się na poziomie ok. 14% wynika z ustalonych dla poszczególnych jednostek maszyny harmonogramów konserwacji okresowej. Czas ten można by nieco zmniejszyć, gdyby Ośrodek posiadał większą liczbę urządzeń zewnętrznych, co umożliwiłoby rozkładanie zabiegów konserwacyjnych w czasie bez potrzeby wstrzymywania normalnego toku pracy maszyny.

To samo dotyczy również przestojów awaryjnych. Techniczna sprawność maszyny, jej pewność działania jest na dobrym poziomie. Mimo to w ciągu czterech pierwszych miesięcy 1968 r. przestoje awaryjne wyniosły 4%. Analiza poszczególnych przestojów wykazuje jednak, że w większości przypadków przestój maszyny spowodowany jest drobnymi na ogół niesprawnościami urządzeń peryferyjnych, głównie czytnika kart. Brak rezerwowych urządzeń peryferyjnych powoduje przy najdrobniejszych niesprawnościach mechanicznych tych urządzeń — przestój całej maszyny.

Takim wąskim gardłem w maszynie ICT-1905, pracującej w GUS jest czytnik kart. W sytuacji, kiedy masowe dane jednostkowe wprowadza się wyłącznie z kart dziurkowanych (a o masowości świadczy fakt, że w okresie od początku roku do 15 maja 1968 — do maszyny wczytano 3845 tys. kart), nie licząc kart z programami, maszyny powinny być bezwzględnie wyposażone minimum w dwa czytniki kart.

Uwagi końcowe

Wykonane za pomocą maszyny elektronicznej opracowania statystyczne wykazały poważne efekty, zarówno pod względem zmniejszenia pracochłonności wykonania, jak i poprawy jakości wyników, dzięki bardziej szczegółowej automatycznej kontroli danych jednostkowych. Maszyna elektroniczna zaczyna także wywierać pozytywny wpływ na szereg innych elementów organizacji badań i opracowań statystycznych. Zagadnienia te jednak z uwagi na ich wagę wymagają odrębnego omówienia.

Z TEZ KC PZPR NA V ZJAZD PARTII

Nieodzownym warunkiem usprawnienia zarządzania przemysłem winno być wzmocnienie sprzężonej z racjonalnym planowaniem kontroli, polegającej na zapewnieniu szybkiej informacji o odchyleniach od planowanego przebiegu prac dla umożliwienia bieżącej korekty działania, a w razie potrzeby — korekty planu. System informacji i statystyki w gospodarce narodowej powinien być dostosowany do współczesnych wymagań systemu zarządzania i nowoczesnej techniki.
