



Grey Scale #13



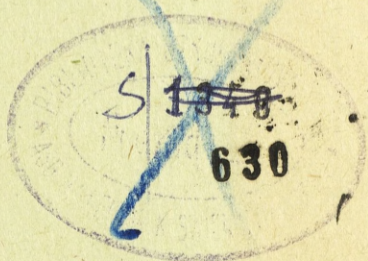
A 1 2 3 4 5 6 M 8 9 10 11 12 13 14 15 B 17 18 19

AKADEMIA SZTABU GENERALNEGO
im. generała broni K. Świerczewskiego

20

pplk mgr inż. Kazimierz ŁĘCKI

**PROBLEMATYKA WYSZUKIWANIA
INFORMACJI**



4191

WARSZAWA

PAŹDZIERNIK

1968

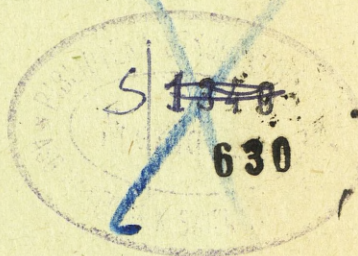


AKADEMIA SZTABU GENERALNEGO
im. generała broni K. Świerczewskiego

20

ppłk mgr inż. Kazimierz ŁECKI

PROBLEMATYKA WYSZUKIWANIA
INFORMACJI



4191

WARSZAWA

PAŹDZIERNIK

1968

AKADEMIA SZTABU GENERALNEGO
im.gen.broni K. Świerczewskiego

20

ppłk mgr inż. Kazimierz IĘCKI

PROBLEMATYKA WYSZUKIWANIA INFORMACJI



WARSZAWA

październik

1968 r.

SPIS TRESCI

	str.
1. Wstęp	1
2. Podział systemów informacyjnych	3
3. Wyszukiwanie informacji	5
4. Kryteria wyszukiwania informacji	12
5. Metody wyszukiwania informacji	13
6. Niezautomatyzowane systemy wyszukiwania informacji	15
7. Ręczne systemy wyszukiwania informacji.....	15
8. Zmechanizowane systemy wyszukiwania informacji ...	22
9. Zautomatyzowane systemy wyszukiwania informacji ..	25
10. Wyszukiwanie informacji przy zlagodzonych kryteriach	30
11. Zakonczenie	33

Cywilizacja rozwijała się dzięki gromadzeniu, przetwarzaniu i wykorzystywaniu informacji. Jednakże od wielu już lat zaistniał dziwny paradoks polegający na tym, że dalszy rozwój cywilizacji zaczyna być hamowany nadmiarem informacji. Dzieje się tak dlatego, ponieważ nagromadzenie w świecie ogromnych zbiorów informacji wywołało trudności w docieraniu do informacji potrzebnych. Stwierdzono np., że liczba wydawnictw naukowych rośnie wykładniczo tzn. udsięciokrotnia się w ciągu każdych 50 lat. Już 100 lat temu pracownik naukowy nie był w stanie przeczytać wszystkich wydawanych w świecie publikacji naukowych z zakresu swej specjalności. Fakt ten, nazwany później przez niektórych historyków nauki pierwszym progiem absurdu, spowodował że zaczęto wydawać czasopisma przeglądowe, stanowiące streszczenia publikacji naukowych. Przeglądy te, śledzone na bieżąco przez zainteresowanych, umożliwiały im docieranie do źródeł potrzebnych informacji. Jednak z biegiem czasu zaczęło pojawiać się tak dużo czasopism przeglądowych, że czytanie ich na bieżąco przekraczało możliwości zainteresowanych. Wprowadzono więc wydawanie przeglądów czasopism przeglądowych, co z kolei nazwane drugim progiem absurdu. Już od kilku lat nadmierna ilość przeglądów z przeglądów utrudnia bardzo, bądź czyni niemożliwym wyszukanie aktualnych informacji.

Obecnie ukazuje się rocznie w świecie około 50 tys. czasopism naukowo - technicznych, 100 tys. książek naukowych, 250 tys. opisów patentowych. Biblioteki świata dysponują 30 milionami tytułów dzieł naukowych a urzędy patentowe 12 milionami opisów patentowych. Ten potop informacji wywołuje zwłoki w uzyskiwaniu informacji potrzebnych, co z kolei prowadzi często do ich dezaktualizacji, do dublowania prac naukowych i przynosi ogromne straty. Szereg autorów prac naukowych wyraża pogląd, że wartość wysiłków, związanych z wyszukiwaniem niezbędnych informacji, stanowi około 50% wartości tych prac. O słuszności tego poglądu świadczą następujące przykłady.

Z danych amerykańskich wynika, że ekspert ustalając, czy dany wynalazek jest nowy, traci ponad 60% czasu tylko na wyszukiwanie informacji w biurach patentowych. Chemicy poświęcają około 50% czasu na wyszukiwanie i ewidencję informacji, a tylko 35% czasu przeznaczają na eksperymenty.

O ujemnych skutkach wywołanych trudnościami w uzyskiwaniu aktualnych informacji świadczą następujące przykłady.

Dublowanie badań naukowych, spowodowane niedocieraniem na czas niezbędnych informacji, przynosi Amerykanom roczne straty przekraczające 2,6 mld. dolarów. Z tych samych powodów Anglicy tracą rocznie 15 mln. funtów. Ze względu na trudności w uzyskiwaniu potrzebnych informacji i związane z tym koszty wielu konstruktorów japońskich rozwiązuje szereg problemów od nowa, wiedząc że problemy te zostały już wcześniej rozwiązane. W tym stanie rzeczy biblioteki naukowe, archiwa, urzędy patentowe i inne instytucje, zajmujące się gromadzeniem, przechowywaniem i wydawaniem informacji, nie mogą w dalszym ciągu stosować tradycyjnych środków i metod wyszukiwania informacji. Środki i metody te, oparte o katalogi, biuletyny informacyjne i poradniki bibliograficzne, charakteryzują się bowiem długim czasem wyszukiwania informacji i nie gwarantują jej wszechstronności, dokładności i aktualności.

Pozostający w ścisłym związku z rozwojem nauki gwałtowny rozwój techniki doprowadził do powstania wielkich przedsiębiorstw, w których procesy produkcyjne zostały w dużym stopniu zautomatyzowane. W tych warunkach skomplikowała się organizacja przedsiębiorstw oraz kierowanie nimi. Wymagają one bowiem kierowania z taką szybkością i dokładnością, która przekracza możliwości człowieka. Typowym problemem takich przedsiębiorstw jest konieczność szybkiego ewidencjonowania, przetwarzania i wyszukiwania coraz to większych ilości informacji. Ponieważ w wielu przypadkach procesy te muszą być realizowane w bardzo krótkim czasie bądź nawet w czasie rzeczywistym, problemu tego nie można więc rozwiązać w drodze zwiększania liczby pracowników administracyjnych.

Analogiczne problemy informacyjne występują w wielu innych dziedzinach. W komunikacji lotniczej niezbędne jest ewidencjonowanie i wydawanie informacji o bieżącym położeniu samolotów w przestrzeni; w meteorologii - o sytuacji atmosferycznej; w handlu - o ruchu towarów na rynku i stanie zapasów. Ewidencjonowanie, przetwarzanie i wydawanie informacji jest konieczne przy sporządzaniu planów gospodarczych, prowadzeniu rachunkowości bankowej, obliczaniu płac.

Problemy informacyjne tego rodzaju posiadają szczególne znaczenie w wojsku. W okresie pokojowym ewidencjonuje się i uaktualnia informacje o potencjale ewentualnych przeciwników; prowadzi się

dokładną ewidencję środków własnych, ludzi podlegających mobilizacji itp. Prowadzenie działań bojowych wiąże się również ściśle z ewidencjonowaniem, przetwarzaniem i przekazywaniem informacji. Dotyczy to danych o nieprzyjacielu jak i o siłach własnych, danych o warunkach atmosferycznych, ewidencji strat i uszkodzeń itp. Jak dalece rozbudowana jest ewidencja wojskowa i jakie wywołuje skutki, świadczą o tym dane amerykańskie, dotyczące sprzętu wojskowego. Z danych tych wynika, że samochód wymaga średnio około 4 tys., a czołg około 6 tys. części zamiennych, że nomenklatura części zamiennych w służbie technicznej i kwatermistrzowskiej obejmuje 200 tys. nazw. Wg tychże danych, podczas II wojny światowej w organach tyłowych tracone 80% czasu na wystawianie i przesyłanie zapotrzebowań i dokumentów materiałowych, a tylko pozostałą część czasu wykorzystywano na wydawanie i dostawy materiałów. W warunkach wojny współczesnej sprawa skomplikuje się jeszcze bardziej, gdyż na planowanie zabezpieczenia materiałowo-technicznego oraz prowadzenie, związanej z tym ewidencji, przeznaczają się coraz to mniej czasu.

Kilka przytoczonych powyżej przykładów wskazuje na to, że istnieje konieczność tworzenia w wielu dziedzinach, w tym i w wojsku, systemów umożliwiających przechowywanie wielkich zbiorów informacji oraz szybkie jej aktualizowanie, wyszukiwanie i przekazywanie. Systemy takie nazywano informacyjnymi. W ostatnich latach obserwuje się bardzo szybki rozwój systemów informacyjnych różnego rodzaju. O dynamice ich rozwoju wydatki przeznaczane na ten cel. Np. w 1960 roku wartość wyprodukowanych urządzeń informacyjnych w USA wynosiła około 3 mln dolarów; w roku 1965 osiągnęła pułap 100 mln. dolarów, a przewiduje się, że w dalszym ciągu wartość ta co 3 lata będzie podwajana.

Obecnie na całym świecie znajdują coraz szersze zastosowanie systemy informacyjne budowane na bazie elektronicznych maszyn cyfrowych. Zgodnie z danymi amerykańskimi na systemy informacyjne tego rodzaju wydano w skali światowej w 1965 r. 6,5 mln dolarów, a w 1969 r. przewiduje się na ten cel 110 mln dolarów. Dane te nie obejmują maszyn stosowanych w wojskowych systemach informacyjnych.

Podział systemów informacyjnych

Wszystkie systemy informacyjne mają przede wszystkim zapewnić wydawanie różnego rodzaju informacji z odpowiednich zbiorów informacji i pracują wg zasady "pytanie - odpowiedź". W zależności od tego, czy

wprowadzana do nich informacja może podlegać przetwarzaniu logicznemu, czy też nie, systemy informacyjne można podzielić na dwie grupy:

- a/ systemy do wyszukiwania informacji;
- b/ -"- informacyjno-logiczne.

Systemy do wyszukiwania informacji charakteryzują się tym, że można z nich otrzymać informację tylko w takiej postaci, w jakiej została do nich wprowadzona.

Systemy informacyjno-logiczne mogą nie tylko wydawać wprowadzoną do nich uprzednie informację, ale umożliwiają także logiczne przetworzenie jej, dzięki czemu mogą wydawać nową informację, jaka do nich nie została wprowadzona.

Obecnie znajdują praktyczne zastosowanie jedynie systemy do wyszukiwania informacji i tylko takie będą w dalszym ciągu rozważane. Systemy informacyjno - logiczne, będące w stadium początkowych eksperymentów, rozwijają się dopiero wówczas, gdy zostanie rozwiązanych szereg problemów logicznych, językowych i technicznych.

Systemy do wyszukiwania informacji można podzielić ze względu na charakter wydawanych informacji na:

- a/ dokumentacyjne;
- b/ faktograficzne.

Systemy dokumentacyjne są to takie systemy, które wydają użytkownikowi oryginały lub kopie, bądź adresy miejsca przechowywania dokumentów zawierających informację, stanowiącą odpowiedź na zadane pytanie.

Systemy faktograficzne wydają użytkownikowi faktyczną informację stanowiącą odpowiedź na zadane pytanie.

Rodzaj stosowanych środków technicznych daje podstawę do podziału systemów do wyszukiwania informacji na:

- a/ ręczne;
- b/ zmechanizowane;
- c/ zautomatyzowane.

Systemy uwzględnione w tym podziale rozwijają się kolejno i różnią się złożonością, możliwościami oraz stopniem udziału człowieka i urządzeń technicznych w procesie wyszukiwania informacji. Zastosowanie spośród nich tego lub innego systemu uwarunkowane jest wieloma czynnikami, a przede wszystkim:

- a/ wymaganą szybkością wyszukiwania informacji;
- b/ wielkością zbiorów informacji;
- c/ opłacalnością.

W przypadku, gdy nie jest wymagany b. krótki czas wyszukiwania informacji, a zbiory nie osiągają zbyt dużych wielkości, to zadowalające wyniki daje wyszukiwanie ręczne.

Przy większych zbiorach i bardziej ograniczonym czasie wyszukiwania może okazać się najbardziej racjonalny system zmechanizowany.

Gdy w grę wchodzi b. duże zbiory i wymagane są b. krótkie czasy wyszukiwania informacji, to konieczne jest zastosowanie systemu zautomatyzowanego. System taki może być również niezbędny i przy mniejszych zbiorach, o ile potrzebny jest natychmiastowy dostęp do informacji.

Przy wyborze rodzaju systemu informacyjnego należy mieć na uwadze wszystkie argumenty "za i przeciw", aby wybrany system spełniał stawiane mu zadania i był opłacalny.

W dalszej części rozpatrzone zostanie tylko zagadnienie wyszukiwania informacji w zbiorach. Warto jednak wspomnieć, że problem ewidencji i przechowywania wielkich zbiorów informacji rozwiązywany jest głównie w oparciu o różne sposoby zapisu elektronicznego. Sposoby te umożliwiają gromadzenie b. dużej ilości informacji za pomocą środków technicznych o niewielkich rozmiarach. Problem zaś szybkiego przekazywania dużych ilości informacji rozwiązywany jest za pomocą technicznych środków transmisji danych.

Wyszukiwanie informacji

Z wyszukiwaniem informacji wiąże się przede wszystkim następująca problematyka:

- a/ język wyszukiwania informacji;
- b/ metody wyszukiwania informacji;
- c/ środki wyszukiwania informacji.

Język wyszukiwania informacji jest sformalizowanym systemem semantycznym, przeznaczonym do symbolicznego wyrażania:

- a/ sensu informacji zawartej w dokumentach;
- b/ sensu pytań.

Dyspensując takim językiem, procedurę wyszukiwania informacji sprowadzamy do porównywania symboli opisujących treść dokumentów z symbolami wyrażającymi treść pytania. W przypadku zgodności tych symboli wynik wyszukiwania jest pozytywny.

Wobec języka do wyszukiwania informacji stawia się następujące podstawowe wymagania:

- a/ język wyszukiwania powinien być jednoznaczny, tzn. powinien dopuszczać jedną i tylko jedną interpretację treści /w związku z tym nie może zawierać synonimów, homonimów i polisemii/;
- b/ język wyszukiwania powinien posiadać dostateczną "siłę semantyczną", by mógł wyrażać w swoich terminach dowolny tekst języka stosowanego w dokumentach i pytaniach.

W praktyce stosowane są 3 typy sformalizowanych języków wyszukiwania informacji:

- a/ języki klasyfikacji bibliograficznej, z których najbardziej rozpowszechnił się język noszący nazwę "uniwersalnej klasyfikacji dziesiętnej";
- b/ języki klasyfikacji przedmiotowej, w których hasła przedmiotowe wykorzystywane są do oznaczania treści dokumentów;
- c/ języki typu deskryptorowego, które stanowią w zasadzie rozwinięcie języków klasyfikacji przedmiotowej.

Obecnie rozpowszechniony jest pogląd, że do mechanizacji i automatyzacji procesów związanych z wyszukiwaniem informacji nadają się najbardziej języki typu deskryptorowego.

Omówienie problematyki związanej z wyszukiwaniem informacji za pomocą języka deskryptorowego wymaga uprzedniego wyjaśnienia kilku pojęć związanych z tym językiem, a mianowicie takich, jak:

- a/ słowo kluczowe;
- b/ deskryptor;
- c/ tezaurus;
- d/ element informacji;
- e/ obraz dokumentu;
- f/ obraz pytania;
- g/ indeksowanie asocjatywne.

Nazwą słowa kluczowe określane są najbardziej charakterystyczne słowa języka naturalnego, występujące w dokumentach i obrazujące ich treść znaczeniową.

Synonimy słów kluczowych i bliskoznacznych łączy się w jedną grupę i zastępuje jednym odpowiednio wybranym słowem kluczowym. Takie słowo kluczowe zwane jest deskryptorem. Na ogół jednak termin deskryptor używany jest w szerszym sensie. W takim przypadku

deskryptorami nazywane są wszelkie słowa kluczowe, zwroty utworzone ze słów kluczonych a nawet pojedyncze znaki, o ile elementy te używane są do opisu treści znaczeniowej dokumentów.

Przykładami deskryptorów mogą być: tytuły zagadnień omawianych w artykule, dane personalne pracownika, dane taktyczno-techniczne sprzętu wojskowego, zakresy wielkości fizycznych aparatury technicznej itp.

Termin tezaurus stosowany jest jako nazwa specjalnego słownika deskryptorów wykorzystywanego do opisywania jakiejś dziedziny nauki.

Poniżej podany jest przykład tezaury służącego do opisywania dziedziny, jaką stanowią maszyny elektryczne /silniki, prądnice, generatory, przetwornice/.

TEZAUZUS
KARTOTEKI "MASZYNY ELEKTRYCZNE"

SILNIKI EL. do 600 W

1. trójfaz. asynchroniczne;
2. jednofaz. z fazą rozruchową;
- - - - -
- - - - -
11. prądu stałego.

MASZYNY EL. POWYŻEJ 600 W

12. silniki dwufaz. asynchroniczne;
13. silniki trójfaz. synchroniczne;
- - - - -
- - - - -
29. kompensator synchroniczny

PRZETWORNICE

30. do przetw. pr. trójfaz. 50 Hz na pr. stały;
31. - " - pr. jednofaz. - " - " -
- - - - -
- - - - -
50. selsymy.

SPECJALNE MASZYNY EL.

51. prądu stałego;
52. -"- zmiennego.

MOC MASZYNY EL.

- 53. do 50 W;
- 54. 51 do 120 W;
-
-
- 80. powyżej 100 000 kW.

NAPIĘCIE MASZYNY EL.

- 81. 6 V;
- 82. 12 V;
-
-
- 95. 13,8 kV i powyżej.

OBROTY MASZYN EL.

- 96. 100 obr/min. i poniżej;
- 97. 101 - 105 obr/min;
-
-
- 114. z szerokim przedziałem regulacji.

CZĘSTOTLIWOŚĆ

- 115. poniżej 50 Hz;
- 116. 50 Hz;
-
-
- 130. 10 000 Hz.

OSŁONA MASZYN EL.

- 131. wodoszczelne;
- 132. odkryte;
-
-
- 138. zakryte przedmuchiwane.

SPECJALNE WYKONANIA MASZYN EL.

- 139. bezszumne;
- 140. z reduktorem;
-
-

143. przystosowane do warunków tropikalnych.

SPOSOB MONTAŻU

144. poziome z kadłubem na łapach;

145. pionowe - " - - " -

146. poziome - " - bez łap;

147. pionowe - " - - " -

MODYFIKACJA SILNIKÓW ASYNCH.

148. z powiększonym momentem rozruchu;

149. - " - poślizgiem;

.....

.....

151. silnik el. dla przemysłu tekstylnego.

ODPORNOŚĆ IZOLACJI NA NAGRZEWANIE

152. do 105°C

153. do 120°C;

157. el. maszyny zdjęte z produkcji.

Tablica

Nr	I lite- ra	Nr	II lite- ra	Nr	III lite- ra	Nr	I cyfra	Nr	II cyfra
158	A	183	A	208	A	233	0	243	0
159	B	184	B	209	B	234	1	244	1
.
.
.
.	242	9	252	9
.
.
.
182	Z	207	Z	232	Z				

Jak widać z przytoczonego słownika, dowolną maszynę elektryczną można opisać za pomocą 252 deskryptorów. Tablica umieszczona na końcu słownika zawiera deskryptory, jakie stanowią litery alfabetu łacińskiego i cyfry systemu dziesiętnego. Znaki te wykorzystywane są do oznaczania serii i typów maszyn.

Budowa tezaury dla dziedziny "maszyny elektryczne" jest stosunkowo prosta, ponieważ język w dziedzinie elektrotechniki, jak i w innych dziedzinach techniki, jest w dużym stopniu sformalizowany.

Opracowanie tezaurusów jest problemem złożonym i wymagającym oddzielnego potraktowania, przekraczającego ramy niniejszego skryptu.

Elementem informacji nazywać będziemy informację określającą pojedynczy dokument /reprezentującą ten dokument/.

Przykładami elementów informacji mogą być: tytuł książki wraz z imieniem i nazwiskiem jej autora oraz datą wydania, imię i nazwisko pracownika, marka sprzętu technicznego itp.

Obrazem dokumentu nazywać będziemy informację wyrażającą za pomocą deskryptorów podstawową treść znaczeniową dokumentu.

Obrazem pytania nazywać będziemy treść pytania wyrażoną za pomocą deskryptorów.

Obrazy dokumentów i pytań tworzone są za pomocą tezaurusów odpowiednich dziedzin.

Indeksowaniem asocjatywnym nazywany jest proces wyrażania treści znaczeniowej dokumentu lub pytania za pomocą języka deskryptorowego.

Indeksowanie asocjatywne wymaga wiernego odtworzenia treści dokumentu i pytania. Dlatego też może być ono pomyślnie realizowane przez pracowników będących specjalistami w dziedzinie, której dotyczy dokument, jak również dobrze zapoznanych z budową tezaury danej dziedziny oraz zasadami indeksowania.

Proces wyszukiwania informacji za pomocą języka deskryptorowego wygodnie jest rozpatrzyć, wykorzystując do tego celu zbiory prostokątne /rys. 1/. Zbiory takie wyrażają zależności zachodzące między elementami informacji a deskryptorami, przy czym zależności te wyrażone są za pomocą zer i jedynek naniesionych na przecięciach kolumn elementów infor -

El. inform. Deskryptory	e_1	e_2	...	e_m
d_1	1	0		1
d_2	0	1		1
⋮				
d_n	1	1		0

Rys. 1

macji i wierszy deskryptorów. Jedynka oznacza, że dany element informacji reprezentuje dokument zawierający dany deskryptor, a zero oznacza, że takiego związku nie ma.

Podzbiór zer i jedynek kolumny stanowi więc obraz dokumentu reprezentowanego przez element informacji tej kolumny.

Zbiory prostokątne mogą być wykorzystywane do wzrokowego wyszukiwania informacji, jak również stanowią materiał wyjściowy do organizacji zbiorów do wyszukiwania informacji za pomocą różnych metod i środków technicznych.

Wyszukiwaniem informacji nazywać będziemy proces ujawniania elementów informacji /a tym samym odpowiadających im dokumentów/ związanych wg określonego kryterium z zadanymi deskryptorami.

Tak rozumiany proces wyszukiwania informacji może być prowadzony w drodze sortowania odpowiednio zorganizowanych zbiorów.

Podstawowe elementy systemu do wyszukiwania informacji

W skład systemu do wyszukiwania informacji wchodzi następujące elementy podstawowe:

- a/ tezaury;
- b/ zbiory informacji;
- c/ środki techniczne do sortowania zbiorów.

Na zbiory informacji składają się zwykle:

- a/ zbiory do wyszukiwania /informacji/;
- b/ zbiory podstawowe.

Zbiór do wyszukiwania /informacji/ zbudowany jest z elementów informacji i deskryptorów /w zbiorze tym prowadzony jest proces wyszukiwania elementów informacji, związanych wg określonego kryterium z grupą deskryptorów zadanych w pytaniu/.

Jego organizacja zależy od przyjętych metod wyszukiwania informacji, co rozpatrzone zostanie w dalszej części skryptu.

W skład zbioru podstawowego wchodzi dokumenty zawierające wszelkie dane /pełną informację/ dotyczące poszczególnych elementów informacji zbioru do wyszukiwania. Dokumentami tymi mogą być: zbiory biblioteczne, archiwalne i patentowe, dokumentacja techniczna, teazki personalne itp. Treść dokumentów zbioru podstawowego może być naniesiona na papierze, na kartach perforowanych, na nośnikach pamięci zewnętrznej elektronicznych maszyn cyfrowych itp.

Kryteria wyszukiwania informacji

Wyszukiwanie informacji może być prowadzone wg różnych kryteriów. Np. wyszukiwanie informacji w zbiorze, zawierającym dane taktyczno-techniczne o sprzęcie pancernym państw NATO, może polegać na ustaleniu nazw wszystkich czołgów, które zostały wyprodukowane w latach 1960-1967, mają kaliber armaty 85 mm lub wyższy, posiadają pancerz o grubości od 45 mm wzwyż i rozwijają prędkość maksymalną ponad 45 km/godz. /w przykładzie tym nazwy czołgów stanowią elementy informacji, a lata 1960-1967, kalibry armat od 85 mm lub wyższe, grubości pancerza powyżej 45 mm oraz prędkości maksymalne od 45 km/godz. wzwyż - deskryptory tych elementów/.

Innym przykładem wyszukiwania informacji w zbiorze, zawierającym dane personalne żołnierzy, może być ujawnianie nazwisk i imion takich oficerów, którzy posiadają stopień majora lub podpułkownika, ukończyli wyższe studia techniczne i dowódcze, znają język angielski i nie zamieszkują w Warszawie.

Kryteria wyszukiwania informacji można formułować zdaniami logicznymi. Rozpatrzmy to na prostym przykładzie ogólnym. W tym celu zajmiemy się zbiorem przedstawionym na rys. 2. W skład tego zbioru wchodzi 6 elementów informacji i 5 deskryptorów. Elementy informacji oznaczone zostały cyframi od 1 do 6, a deskryptory - kolejnymi literami alfabetu od A do E.

Elementy informacji	1	2	3	4	5	6
Deskryptory						
A	1	1	0	0	0	1
B	1	1	0	0	1	1
C	0	0	1	0	0	1
D	1	1	0	1	0	1
E	1	0	1	1	1	0

Rys. 2

W zbiorze tym prowadzić będziemy przykładowo wyszukiwanie informacji wg następujących kryteriów:

- a/ ustalić wszystkie elementy informacji, które zawierają jednocześnie deskryptory B i D;

b/ ustalić wszystkie elementy informacji, które zawierają deskryptory B i C lub E/.

W celu zrealizowania pierwszego wyszukiwania tworzymy iloczyn logiczny wierszy B i D.

	1	2	3	4	5	6
B	1	1	0	0	1	1
D	1	1	0	1	0	1
B i D	1	1	0	0	0	1

Te elementy informacji, dla których iloczyn logiczny wynosi 1, stanowią wyniki wyszukiwania, a więc elementy: 1, 2 i 6.

Aby zrealizować drugie wyszukiwanie, tworzymy najpierw sumę logiczną wierszy C i E.

	1	2	3	4	5	6
C	0	0	1	0	0	1
E	1	0	1	1	1	0
C lub E	1	0	1	1	1	1

Z kolei tworzymy iloczyn logiczny: B i /C lub E/.

	1	2	3	4	5	6
B	1	1	0	0	1	1
C lub E	1	0	1	1	1	1
B i /C lub E/	1	0	0	0	1	1

Wynikami wyszukiwania wg ostatniego kryterium są więc elementy: 1, 5, 6.

Za pomocą algebry Boole'a można określać bardziej złożone kryteria wyszukiwania informacji, niż przytoczone powyżej.

Metody wyszukiwania informacji

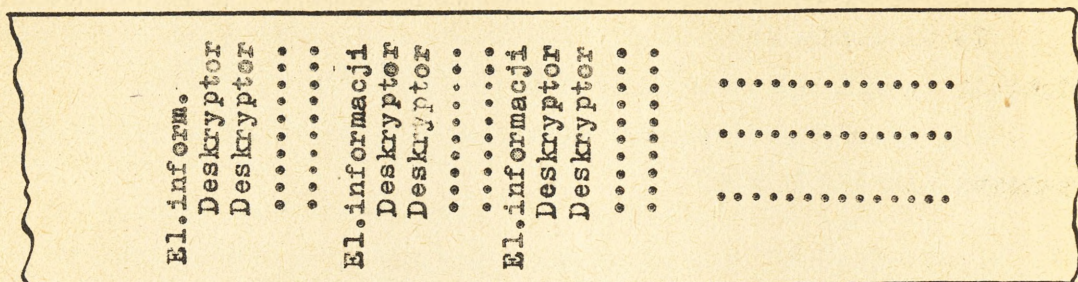
Do wyszukiwania informacji w zbiorach można stosować trzy metody:

- a/ bezpośrednią;
- b/ inwersyjną;
- c/ tablicową.

Każda z tych metod charakteryzuje się odmienną organizacją zbioru informacji.

Występujące w zbiorach elementy informacji i deskryptory koduje się liczbowo. Kodowanie ma na celu zmniejszenie rozmiarów zbiorów i ułatwienie procesu wyszukiwania informacji.

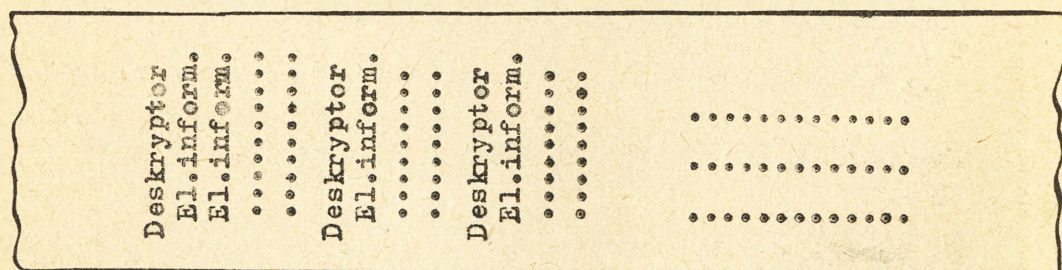
W przypadku metody bezpośredniej zbiorów do wyszukiwania zorganizowany jest w postaci spisu elementów informacji i związanych z nimi deskryptorów tak, jak to wyjaśnia rys. 3.



Rys. 3.

Wyszukiwanie informacji prowadzone metodą bezpośrednią nie wymaga porządkowania zbioru.

Przy stosowaniu metody inwersyjnej zbiorów do wyszukiwania zorganizowany jest wg odwrotnej zasady /w porównaniu z metodą bezpośrednią/, co objaśnia rys. 4.

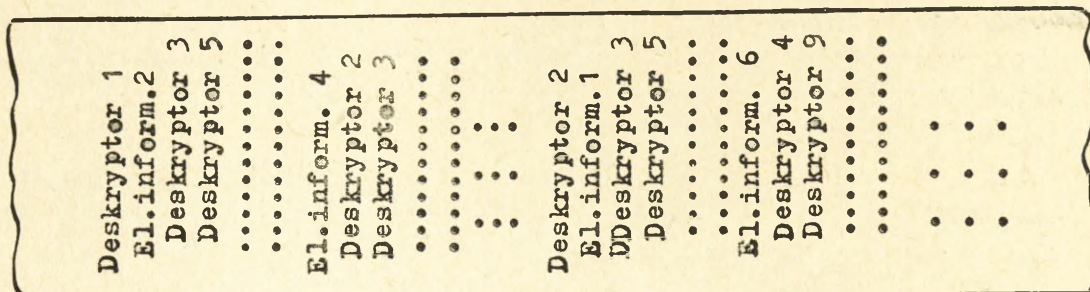


Rys. 4

Przy tej metodzie wyszukiwania informacji wymagane jest uporządkowanie zbioru zgodnie z kolejnością wzrastania kodów /liczbowa lub alfabetyczna/.

Zbiór dla metody tablicowej zorganizowany jest w postaci tablic, tworzonych dla każdego deskryptora. Poszczególne tablice w zbiorze uporządkowane są wg rosnących kodów deskryptorów. Pod deskryptorem każdej tablicy naniesiony jest spis związanych z nim elementów informacji. Przy każdym elemencie informacji podawane są t

wszystkie deskryptory, które są z nim związane, lecz których kody mają wyższą numerację od kodu deskryptora tablicy. Budowę takiego zbioru ilustruje przykładowo rys. 5.



Rys. 5.

Wyszukiwanie informacji trzema przytoczonymi metodami, przy użyciu różnych środków technicznych, omówione zostanie na przykładzie zbioru podanego na rys. 2. Zajmiemy się tylko wyszukiwaniem polegającym na znalezieniu wszystkich elementów informacji związanych jednocześnie zadaną grupą deskryptorów. Wszelkie inne kryteria wyszukiwania realizowane są analogicznie.

Niezautomatyzowane systemy wyszukiwania informacji

Zautomatyzowane systemy wyszukiwania informacji rozwinęły się w oparciu o systemy niezautomatyzowane. Zasady organizacji zbiorów do wyszukiwania oraz metody wyszukiwania informacji są we wszystkich trzech rodzajach systemów /ręcznych, zmechanizowanych i zautomatyzowanych/ bardzo podobne.

Ręczne systemy wyszukiwania informacji

W systemach tych, zaliczanych do minimalnej mechanizacji wyszukiwania informacji, urządzenia techniczne służą w zasadzie tylko do przechowywania zbiorów informacji, a wszelkie procesy związane z działaniem systemu wykonywane są ręcznie.

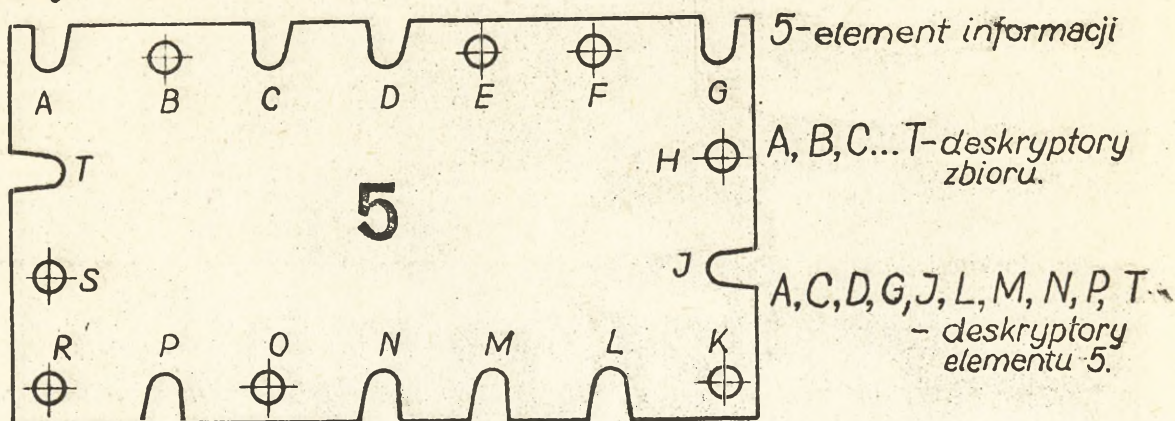
Metoda bezpośrednia

W metodzie bezpośredniej stosowane są karty perforowane dwóch rodzajów:

- a/ karty z perforacją brzegową;
- b/ karty szczelinowe /z perforacją szczelinową/.

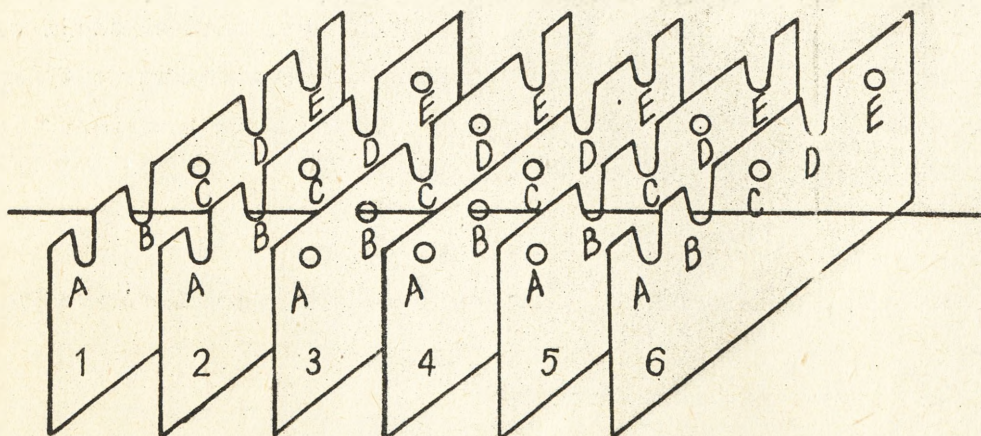
Karty z perforacją brzegową są to karty prostokątne, wokół brzegów których znajdują się otwory kalibrowane, rozmieszczone w jednym lub kilku rzędach. Otwory te stanowią pole kodowe. Każda karta w pliku, stanowiącym zbiór informacji, reprezentuje jeden element informacji, a

otwory brzegowe - deskryptory związane ze zbiorem. Każdemu deskryptorowi odpowiada na każdej karcie otwór wykonany na tej samej pozycji. Wycięcia między otworami a brzegiem karty stanowią kod do oznaczenia deskryptorów związanych z elementem informacji tej karty /rys. 6/, czyli stanowią zakodowany obraz dokumentu.



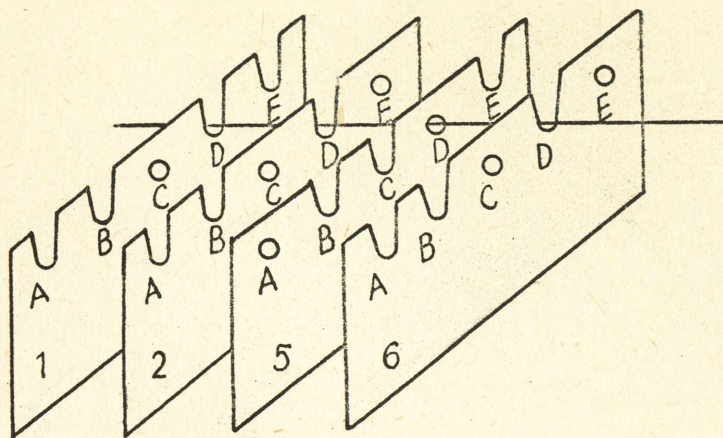
Rys. 6.

Wyszukiwanie informacji, polegające na wysortowaniu ze zbioru tych kart, które odpowiadają poszukiwanym elementom informacji, prowadzone jest za pomocą iglicy. W tym celu przesuwana się ręcznie iglicą na wskroś zbioru kart przez te otwory, które odpowiadają jednemu z zadanych deskryptorów i oddziela ze zbioru te karty, które związane są z wyszukiwaną informacją /karty nie związane z wyszukiwaną informacją pozostają na iglicy/. Rys. 7 przedstawia zasadę wysortowania ze zbioru kart, które posiadają wspólny deskryptor B, tzn. karty z elementami informacji: 1, 2, 5, 6.



Rys. 7.

Aby z kart tych oddzielić karty o innym wspólnym deskrytorze, należy powtórzyć poprzednią czynność. Ilustruje to rys. 8 w odniesieniu do deskrytora D.



Rys. 8.

Z deskrytorami B i D są więc związane elementy informacji 1, 2, 6.

W realizowanej w ten sposób metodzie należy powtórzyć czynność sortowania tyle razy, ile zadanych zostanie deskrytorów. Na wyszukaniu wszystkich elementów informacji, związanych z wszystkimi deskrytorami zadanymi w pytaniu, kończy się proces wyszukiwania faktograficznego. Metoda bezpośrednia pozwala w wielu przypadkach realizować proces wyszukiwania dokumentacyjnego jednocześnie z procesem wyszukiwania faktograficznego. Jest to możliwe wtedy, gdy informację zbioru podstawowego można zmieścić na kartach zbioru do wyszukiwania. W takich przypadkach karty z perforacją brzegową zaopatrzone są w dokumenty zawierające pełną informację o poszukiwanych elementach. Mogą to być przytwierdzone do kart dokumenty oryginalne bądź ich kopie. Przy dużej objętości dokumentu stosowane są mikrokopie, które mogą być wykonane bezpośrednio na karcie albo w postaci mikrofilmu wmontowanego w specjalne okienko karty perforowanej. Ostatnie karty zwane są aperturowymi.

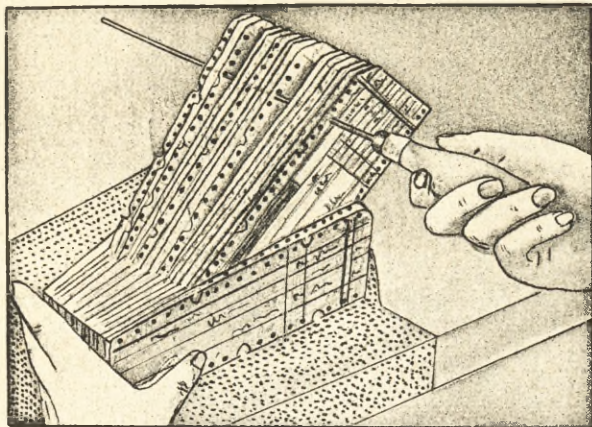
Karty aperturowe są bardzo wygodne, gdyż dokumentacja sporządzona na nich zajmuje około 80% mniej miejsca niż oryginalna. Karty te nadają się szczególnie do przechowywania i wyszukiwania rysunków technicznych, co nabiera dużego znaczenia przy konstrukcjach maszyn złożonych.

Wyszukiwanie informacji za pomocą kart z perforacją brzegową może być realizowane dwoma sposobami:

a/ wyłącznie ręcznie;

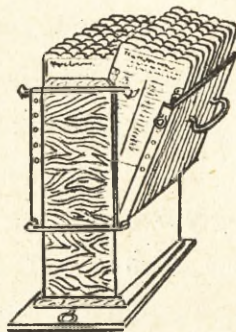
b/ za pomocą półautomatycznego selektora.

Wyszukiwanie wyłącznie ręczne zilustrowane jest na rys. 9.



Rys. 9.

Przykład selektora z odchylaną ścianą boczną pokazany jest na rys. 10.



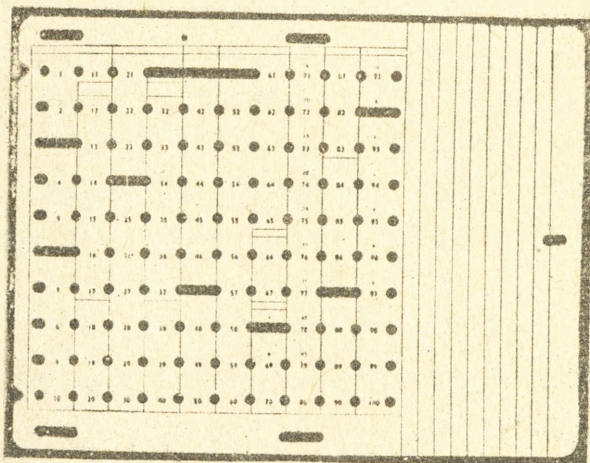
Rys. 10.

Metoda bezpośrednia nie wymaga porządkowania kart w zbiorze, co jest jej zaletą. Perforacja brzegowa ogranicza jednak liczbę deskryptorów, w związku z czym metoda bezpośrednia nadaje się do organizacji wyszukiwania informacji w wąskiej tematyce. Praktycznie można prowadzić tą metodą wyszukiwanie informacji w zbiorach wynoszących do 10 tys. kart - przy sortowaniu za pomocą iglicy, oraz w zbiorach o pojemności do 30 tys. kart - przy wykorzystaniu selektorów. Najczęściej spotyka się zbiory od 2 tys. do 3 tys. kart.

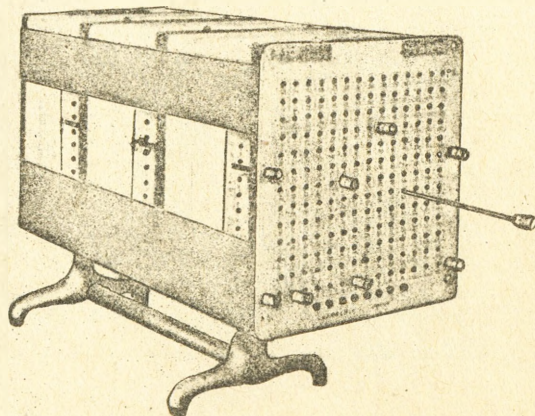
Karty szczelinowe /rys. 11/ posiadają pole kodowe w postaci macierzy, zbudowanej z kalibrowanych otworów, rozmieszczonych w stałej odległości między sobą.

Podobnie, jak w przypadku perforacji brzegowej, każda karta stanowi element informacji, a deskryptory związane ze zbiorem reprezentowane są przez odcinki zawarte między otworami. Mogą to być odcinki poziome albo pionowe /zależnie od przyjętej zasady/. Deskryptory związane z danym elementem informacji zaznaczone są za pomocą szczelin wyciętych między otworami. Wolne pole na kartach wykorzystywane jest na umieszczenie dokumentów bądź ich kopii, które mogą przyjmować takie same formy, jak na kartach z perforacją brzegową.

Wyszukiwanie informacji w zbiorze zbudowanym z kart szczelinowych prowadzone jest za pomocą odpowiedniego selektora. Przykład selektora przedstawia rys. 12. W ścianie bocznej selektora wykonana jest identyczna matryca z otworami, jak na każdej karcie. W celu wyszukania informacji należy przez te otwory selektora, które odpowiadają zadanym deskryptorom,



Rys. 11.



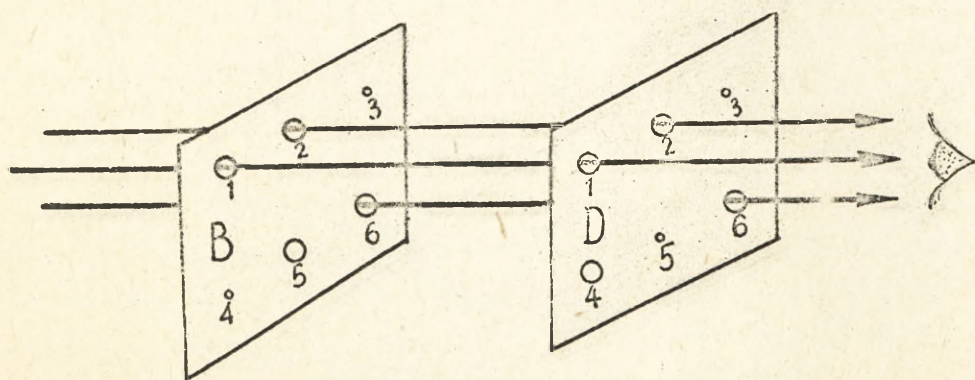
Rys. 12.

przesunąć na wskroś zbioru iglice zwane podstawowymi. Po obróceniu selektora o 90° wszystkie karty z sadanymi deskryptorami wysuną się ze zbioru na odległość równą długości szczelin. Z kolei ustawia się w specjalny otwór wykonany na przeciwległym brzegu ścianki bocznej, iglicę dodatkową i obraca selektor do położenia normalnego. Iglica dodatkowa zapobiega cofnięciu się wysortowanych kart, które, po wyjęciu wszystkich podstawowych iglic, łatwo można oddzielić od zbioru.

Wyszukiwanie informacji za pomocą kart szczelinowych charakteryzuje się podobnymi własnościami, jak wyszukiwanie oparte na kartach z perforacją brzegową.

Metoda inwersyjna

W metodzie bezpośredniej na każdy element informacji zakłada się oddzielną kartę, na której zakodowane są deskryptory, odpowiadające temu elementowi. W metodzie inwersyjnej zbiór na kartach, zwanych superpozycyjnymi, organizowany jest wg odwrotnej zasady. Oddzielną kartę zakłada się na każdy deskryptor. Na każdej karcie koduje się wszystkie elementy informacji, związane z deskryptorem tej karty. W związku z tym ilość kart w pliku, stanowiącym zbiór informacji, odpowiada liczbie deskryptorów zbioru. Elementy informacji kodowane są na kartach za pomocą otworów. Danemu elementowi informacji odpowiada na każdej karcie otwór wykonany na tej samej pozycji. Aby zrealizować wyszukanie informacji dla zadanej grupy deskryptorów, należy wybrać ze zbioru karty odpowiadające tym deskryptorom, nałożyć je dokładnie na siebie /superpozycja/ i obserwować pod światło. Te miejsca, które przeswituje, wskazują wyszukiwane elementy informacji. Zasadę wyszukiwania informacji tą metodą dla zadania rozpatrywanego w metodzie bezpośredniej objaśnia rys. 13.



Rys. 13.

Z rysunku tego widać, że deskryptorem B i D odpowiadają elementy informacji 1, 2, 6.

Celem ułatwienia wybrania kart z niezbędnymi deskryptorami zbiór kart powinien być uporządkowany.

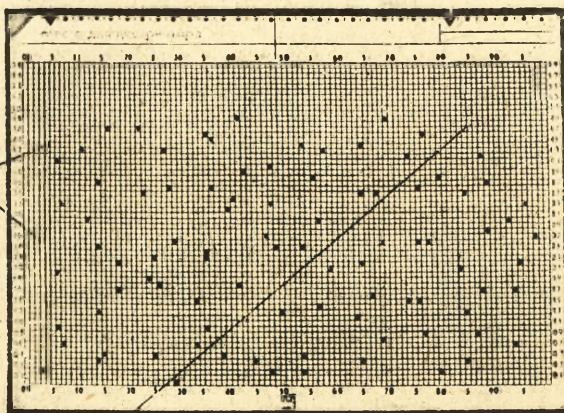
Rozpatrywany system jest systemem faktograficznym, ponieważ za jego pomocą można ustalić tylko "fakty", jakie elementy informacji odpowiadają danym deskryptorom.

W miarę potrzeby pełną informację o wyszukanych elementach informacji można zaczerpnąć ze zbioru podstawowego.

Metoda inwersyjna pozwala organizować wyszukiwanie informacji w tematyce wymagającej dowolnej liczby deskryptorów. Liczba elementów informacji uwarunkowana jest typem kart. Np. stosując w tym celu zwykłe 80-kolumnowe karty perforowane, używane w maszynach analityczno-liczących i w elektronicznych maszynach cyfrowych, można operować liczbą elementów informacji wynoszącą 960. Wykonanie takich kart jest wygodne, ponieważ można je perforować na uniwersalnych dziurkarkach kart a ponadto realizować za ich pomocą wyszukiwanie zarówno ręczne, jak i mechaniczne. Wadą tych kart jest stosunkowo niewielka liczba elementów informacji.

Ogólnie rzecz biorąc, liczba elementów informacji, jaką można zmieścić na jednej karcie superpozycyjnej, jest wprost proporcjonalna do wielkości powierzchni karty i gęstości rozmieszczenia otworów, a odwrotnie proporcjonalna do wielkości powierzchni otworów. Zwiększenie liczby otworów poprzez ich zagęszczanie i zmniejszenie utrudnia perforację oraz wyszukiwanie informacji. Z tego też względu "pojemność karty" podwyższa się zwykle poprzez zwiększanie jej wymiarów. Tą drogą osiągnięto "pojemność karty" wynoszącą nawet 40 tys. elementów informacji. W tym przypadku każdy element informacji oznaczony jest za pomocą współrzędnych /rys. 14/.

otwory



El. inf. 4585

Rys. 14.

Praktycznie metoda inwersyjna stosowana jest do zbiorów zawierających kilkaset deskryptorów i do 10 tys. elementów informacji.

Metoda tablicowa

W tej metodzie używane są karty nieperforowane. Ich liczba w zbiorze równa jest liczbie deskryptorów, gdyż każda karta reprezentuje jeden deskryptor. Na każdej karcie naniesiona jest tablica wykonana wg zasady omówionej już poprzednio. W przypadku przykładu podanego w poprzednich metodach tablica dla deskryptora B miałaby postać przytoczoną na rys. 15.

B		
1		D E
2		D
5		C E
6		D

Rys. 15

Aby ułatwić wyszukiwanie, należy karty ułożyć zgodnie z kolejnością porządkową deskryptorów. W takiej też kolejności muszą być zadane deskryptory. Jeżeli chcemy tą metodą posłużyć się do znalezienia elementów informacji, związanych z deskryptorami B i D, to odszukujemy najpierw tablicę deskryptora B /przytoczoną na rys. 12/, a z kolei te elementy informacji, które związane są ponadto z deskryptorem D /objęte linią/. Otrzymana odpowiedź daje wynik zgodny z poprzednimi.

Metoda tablicowa jest stosunkowo wygodna i szybka, jednak sporządzanie tablic dla dużych zbiorów jest procesem bardziej złożonym i pracochłonnym niż w przypadku metod poprzednich.

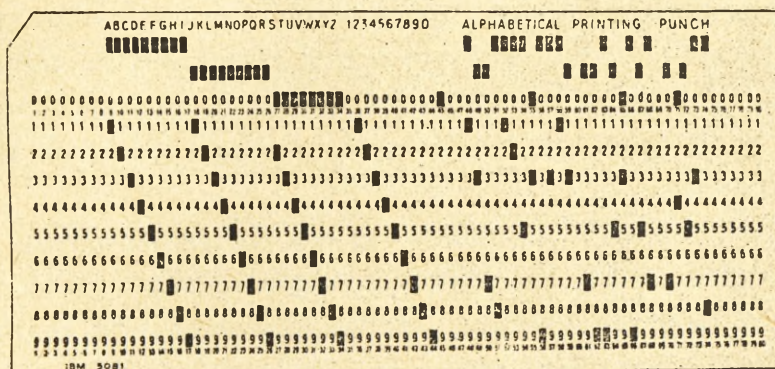
Ze względów praktycznych uporządkowany zbiór kart może być zszyty na wzór książki.

Zmechanizowane systemy wyszukiwania informacji

Zmechanizowane systemy do wyszukiwania informacji organizowane są na bazie maszyn analityczno-liczących. W systemach tych podlegają mechanizacji takie operacje, jak: nanoszenie informacji na nośniki, wyszukiwanie informacji w zbiorze oraz doprowadzanie wyników wyszuki-

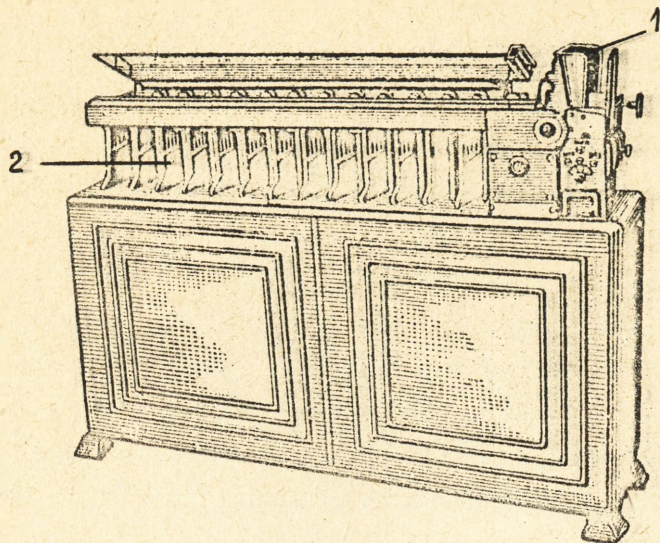
wania do postaci wygodnej dla użytkownika. Pozostałe operacje, niezbędne do działania systemu, wykonywane są ręcznie. Jako nośniki informacji stosuje się karty perforowane najczęściej 80-kolumnowe, rzadziej 90-kolumnowe. Na karty te informacja nanoszona jest za pomocą perforatora. Stosowane są również karty aperturone.

Na rys. 16 przytoczona jest 80-ciokolumnowa karta perforowana z naniesionym na niej kodem cyfr dziesiętnych i liter alfabetu łacińskiego, stosowanym przez amerykańską firmę IBM. Jak widać z rysunku, dowolna cyfra kodowana jest za pomocą jednego otworu, a dowolna litera - za pomocą dwóch otworów. Z prawej strony karty naniesiony jest kod napisu "ALPHABETICAL PRINTING PUNCH". Otwory perforowane są w 12-tu wierszach, z których 10 jest numerowanych cyframi od 0 do 9 a dwa górne nie są oznaczone. Karty zbioru, w którym prowadzony jest proces wyszukiwania informacji, zawierają kody elementów informacji i związanych z nimi deskryptorów.



Rys. 16.

Wyszukiwanie informacji realizowane jest przy użyciu sorterów elektromechanicznych, a wyniki wyszukiwania mogą być wydawane w postaci drukowanej przez tabulatory. Rys. 17 przedstawia wygląd zewnętrzny sortera uniwersalnego. Karty zbioru podlegającego sortowaniu umieszczają się w zasobniku /1/. Mogą być one rozsortowane na 13 grup, zawierających określone cechy, w związku z czym sorter wyposażony jest w 13 kaset odbiorczych /2/. W jednym przebiegu proces sortowania prowadzony jest ze względu na jedną kolumnę. 12 kaset odbiorczych odpowiada 12 wierszom jednej kolumny, a 13 kaset przeznaczona jest na karty nie posiadające otworów w danej kolumnie.



Rys. 17.

Aby znaleźć informację w zbiorze kart należy ten zbiór, lub jego część, umieścić w zasobniku sortera. Sorter wybiera po kolei poszczególne karty i porównuje kody ich deskryptorów z kodami deskryptorów zadanych w pytaniu. W przypadku zaistnienia zgodności tych kodów sorter oddziela daną kartę od zbioru i umieszcza w przeznaczony do tego celu kasecie odbiorczej.

Do wyszukiwania informacji za pomocą systemu zmechanizowanego stosuje się najczęściej metodę bezpośrednią, rzadziej inwersyjną.

Celem ilustracji zmechanizowanego wyszukiwania informacji rozpatrzony zostanie poniżej francuski system "D3D".

W systemie tym zastosowane zostały karty 80-kolumnowe, sortowane z szybkością 700kart/min. Pole karty podzielone jest na dwie części.

Kolumny od 1 do 8 przeznaczone są na zapis kodu elementu informacji, a kolumny pozostałe /9 - 80/ wykorzystane są do zapisu kodów deskryptorów, które mogą być rozmieszczone w dowolnym porządku. System prowadzi więc wyszukiwanie metodą bezpośrednią. Na kod każdego deskryptora przeznacza się 3 kolumny, w związku z czym na karcie mogą być zapisane 24 deskryptory. Do kodowania deskryptorów używane są cyfry od 0 do 9 oraz litery alfabetu łacińskiego, co umożliwia zapis 46655 różnych deskryptorów.

System "D3D" różni się tym od zwykłego sortera, że posiada przystawkę wyposażoną w pamięć i niezbędne obwody logiczne. Dzięki przystawce za jednym obiegiem /sortowaniem/ zostają wydzielone ze zbioru wszystkie karty zawierające zadany deskryptor, bez względu na to, w której kolumnie jest on umieszczony. Po oddzieleniu kart, zawierają-

cych jeden z zadanych deskryptorów, poddaje się je dalszemu sortowaniu ze względu na drugi deskryptor. Z kolei wyselekcjonowane karty z dwoma zadanymi deskryptorami podane zostają sortowaniu ze względu na 3 deskryptor itd.

Wydajność rozpatrywanego systemu charakteryzują następujące dane. Przy zbiorze zawierającym 10 tys. kart, wydzielenie kart z jednym zadanym deskryptorem trwa średnio 15 min. Sortowanie ze względu na drugi zadany deskryptor wymaga około 3 min, ze względu na trzeci - 1 min. itd. W systemie "D3D" stosowane są także karty aperturowe.

Zautomatyzowane systemy do wyszukiwania informacji

Zautomatyzowane systemy do wyszukiwania informacji, w których niemal wszystkie operacje wykonywane są automatycznie, budowane są na bazie elektronicznych maszyn cyfrowych. Zastosowanie tych maszyn do wyszukiwania informacji zapewnia:

- a/ ułatwienie procesu wyszukiwania informacji;
- b/ przyspieszenie - " - - " - - " - -
- c/ zmniejszenie rozmiarów zbiorów informacji;
- d/ usuwanie z nośników informacji nieaktualnych i nanoszenie na ich miejsce nowych.

Elektroniczne maszyny cyfrowe stosowane do wyszukiwania informacji różnią się od uniwersalnych maszyn cyfrowych do obliczeń numerycznych specyfiką operacji, spośród których należy przede wszystkim wymienić:

- a/ wprowadzanie do maszyny, przechowywanie w niej i wyprowadzanie bardzo dużych zbiorów informacji;
- b/ konieczność porządkowania wprowadzanych zbiorów informacji;
- c/ wykonywanie dużej liczby operacji logicznych /szczególnie porównywania/ w stosunku do ilości operacji arytmetycznych;
- d/ działanie na informacji zawierającej słowa zmiennej długości;
- e/ grupowe wprowadzanie i wyprowadzanie informacji.

Przytoczone operacje są typowe dla maszyn do elektronicznego przetwarzania danych. Z tych też względów zadania, związane z wyszukiwaniem informacji, rozwiązywane są zwykle za pomocą systemów elektronicznego przetwarzania danych, o czym świadczy szereg przykładów poda-

wanych w literaturze. Specjalizowane EMC do wyszukiwania informacji spotykane są rzadko.

W przypadku zautomatyzowanego systemu wyszukiwania informacji zbiory informacji przechowywane są w pamięci zewnętrznej, dokąd wprowadzane są z kart perforowanych lub z taśmy perforowanej. W charakterze pamięci zewnętrznej najczęściej stosowana jest taśma magnetyczna, a w maszynach bardziej nowoczesnych - dyski magnetyczne bądź karty magnetyczne. Rodzaj zastosowanej pamięci zewnętrznej w istotny sposób wpływa na szybkość wyszukiwania informacji, która zależy od średniego czasu dostępu do informacji. Czas ten dla taśmy magnetycznej jest rzędu minut, a dla dysków magnetycznych i kart magnetycznych - rzędu msek.

W systemach zautomatyzowanych stosowane są te same metody wyszukiwania informacji, jak w systemach ręcznych. Rozpatrzmy, jak metody te realizowane są w przypadku, gdy pamięć zewnętrzna wykonana jest na taśmie magnetycznej, chociaż te same zasady stosowane są również przy innych rodzajach pamięci.

Metoda bezpośrednia

Przy tej metodzie zbiór na taśmie naniesiony jest w postaci spisu elementów informacji i odpowiadających im deskryptorów, jak to zilustrowane zostało na rys. 3.

Pytanie zawierające spis deskryptorów, dla których należy znaleźć wszystkie związane z nimi elementy informacji, wprowadzone zostaje do pamięci operacyjnej poprzez urządzenie wejściowe bądź za pomocą pulpitu operatora. Proces wyszukiwania informacji przebiega następująco:

Maszyna przepisuje z taśmy magnetycznej do pamięci operacyjnej pierwszy napotkany element informacji wraz z jego deskryptorami. Z kolei następuje porównywanie deskryptorów tego elementu z grupą deskryptorów zadanych. Jeżeli wszystkie deskryptory zadane znajdują się wśród deskryptorów pierwszego elementu, to element ten zostaje zapamiętany w pamięci operacyjnej w miejscu przewidzianym na wynik. Tak postępuje maszyna z każdym kolejnym elementem. W ten sposób po jednokrotnym przewinięciu taśmy magnetycznej, maszyna zarejestruje w pamięci operacyjnej wszystkie elementy informacji, związane z zadaną grupą deskryptorów. Dla metody tej charakterystyczne jest to, że:

- a/ maszyna przenosi z taśmy magnetycznej do pamięci deskryptory wszystkich elementów informacji i dokonuje porównywania ich z deskryptorami zadanymi;

b/ przy zapisywaniu zbioru na taśmie można stosować dowolną kolejność elementów informacji.

Pierwsza z przytoczonych właściwości jest wadą, gdyż na proces wyszukiwania informacji traci się dużo czasu maszynowego. Druga właściwość stanowi zaletę, ponieważ zapisany na taśmie zbiór nie wymaga uprzedniego porządkowania.

Metoda inwersyjna

W przypadku tej metody zbiór na taśmie naniesiony jest w postaci spisu deskryptorów, kolejno uporządkowanych, i odpowiadających im elementów informacji, jak to objaśnione zostało na rys. 4.

Grupa deskryptorów zadanych w pytaniu podlega w pamięci operacyjnej takiemu samemu uporządkowaniu.

Proces wyszukiwania informacji jest następujący.

Maszyna przegląda na taśmie magnetycznej spis deskryptorów i, napotkawszy pierwszy z zadanych, przepisuje go do pamięci operacyjnej łącznie z jego elementami informacji i zapamiętuje. Po znalezieniu na taśmie drugiego z zadanych deskryptorów, maszyna przepisuje do pamięci operacyjnej tylko te jego elementy informacji, które były związane z pierwszym deskryptorem i zapamiętuje je. Przy trzecim deskryptorze maszyna przepisuje do pamięci operacyjnej i zapamiętuje tylko te jego elementy, które były związane z dwoma pierwszymi deskryptorami itd. Jeśli maszyna napotka na taśmie strefę nie związaną z zadanymi w pytaniu deskryptorami, nie rozpatruje jej. Tytułem tego zaoszczędza się czas maszynowy, ale metoda ta wymaga uprzedniego porządkowania zbioru i zadanej grupy deskryptorów.

Metoda tablicowa

Organizacja zbioru do wyszukiwania dla metody tablicowej objaśniona została na rys. 5. Uporządkowaniu tablic tego zbioru na taśmie magnetycznej musi odpowiadać takie samo uporządkowanie zbioru deskryptorów zawartych w pytaniu, zapisanym w pamięci operacyjnej.

Wyszukiwanie informacji ma przebieg następujący.

Maszyna szuka deskryptora /na taśmie/ zgodnego z pierwszym zadanym. Jeśli tablica odszukanego deskryptora zawiera jednocześnie pozostałe deskryptory zadane, to odpowiadające im elementy informacji dają wynik wyszukiwania i podlegają zapisowi i zapamiętaniu w pamięci operacyjnej.

Przy metodzie tablicowej maszyna przegląda dokładnie tylko jedną strefę pamięci taśmowej a mianowicie tę, którą stanowi poszukiwana tablica. Metoda ta jest więc najszybsza.

Opisane powyżej trzy metody wyszukiwania informacji za pomocą elektronicznej maszyny cyfrowej zilustrowane zostaną na tym samym przykładzie, który rozpatrywany był przy opisie metod wyszukiwania ręcznego. Zasady zapisu zbiorów na taśmie magnetycznej i wyniki wyszukiwania zestawione są w tabeli nr 2.

Tabela nr 2

Metoda bezpośrednia	Metoda inwersyjna	Metoda tablicowa
1 $\begin{matrix} A \\ \textcircled{B} \\ \textcircled{D} \\ E \end{matrix}$ 2 $\begin{matrix} A \\ \textcircled{B} \\ \textcircled{D} \end{matrix}$ 3 $\begin{matrix} C \\ E \end{matrix}$ 4 $\begin{matrix} D \\ E \end{matrix}$ 5 $\begin{matrix} B \\ E \end{matrix}$ 6 $\begin{matrix} A \\ \textcircled{B} \\ \textcircled{C} \\ D \end{matrix}$	A $\begin{matrix} 1 \\ 2 \\ 6 \end{matrix}$ $\textcircled{B} \begin{matrix} 1 \\ 2 \\ 5 \\ 6 \end{matrix}$ C $\begin{matrix} 3 \\ 6 \end{matrix}$ $\textcircled{D} \begin{matrix} 1 \\ 2 \\ 4 \\ 6 \end{matrix}$ E $\begin{matrix} 1 \\ 3 \\ 4 \\ 5 \end{matrix}$	A $\begin{matrix} 1 \\ B \\ D \\ E \\ 2 \\ B \\ D \\ 6 \\ B \\ C \\ D \\ B \\ \textcircled{1} \\ D \\ E \\ \textcircled{2} \\ D \\ 5 \\ E \\ \textcircled{6} \\ B \\ C \\ D \\ C \\ 3 \\ E \\ 6 \\ D \\ D \\ 1 \\ E \\ 2 \\ 4 \\ E \\ 6 \end{matrix}$
Deskryptory zadane: B, D. Wynik: 1,2,6.	Deskryptory zadane: B, D. Wynik: 1,2,6.	Deskryptory zadane: B, D. Wynik: 1,2,6.

Rozpatrzmy obecnie dwa przykłady zastosowania zautomatyzowanych systemów do wyszukiwania informacji.

Przykład 1.

Najbardziej rozpowszechnionymi systemami informacyjnymi, w sensie rozumianym w niniejszym artykule, są systemy do wyszukiwania informacji bibliograficznej. Wynika to z ogromnego zapotrzebowania na takie systemy. Obok systemów, przeznaczonych wyłącznie do tych celów, stosowane są do wyszukiwania informacji bibliograficznej także uniwersalne systemy elektronicznego przetwarzania danych. Systemy takie prowadzą wyszukiwanie informacji obok innych zadań elektronicznego przetwarzania danych.

Przytoczony poniżej przykład dotyczy modelu systemu do wyszukiwania informacji, który był eksperymentowany w 1962 roku w Darmsztacie. W systemie tym wykorzystana została elektroniczna maszyna cyfrowa IBM 704. W eksperymencie oparto się na bibliotecznym katalogu Ośrodka Atomowego NRF. Katalog ten zawierał 2 tys. tytułów publikacji naukowych z adnotacjami, w jakich bibliotekach znajdują się te publikacje. Na podstawie katalogu opracowany został zbiór publikacji, który naniesiono na 80-kolumnowe karty perforowane. Każda publikacja zbioru opisana została 6 charakterystykami. Pierwszą charakterystykę stanowił numer porządkowy. Druga charakterystyka zawierała do 11 deskryptorów, informujących o treści publikacji oraz do 11 deskryptorów wyrażających jej formę. W trzeciej charakterystyce zawarte były formalne indeksy, w czwartej - rubryka z danymi wyjściowymi, a w piątej - adres miejsca, gdzie znajduje się dana pozycja. Charakterystyka ostatnia przeznaczona była na ewentualne dodatkowe zapisy. Z reguły informacje zawarte w jednej charakterystyce nie mieściły się na jednej karcie perforowanej i były nanoszone na kilka kart /maksimum do 10/.

Zbiór kart perforowanych podlegał przeniesieniu na taśmy magnetyczne, na które zapisywany był za pomocą czytnika perferyjnego, bez udziału jednostki centralnej.

Program opracowany w języku FORTRAN zapewniał wyszukiwanie informacji dla dowolnie zadanej kombinacji deskryptorów. Każde pytanie formułowane było za pomocą deskryptorów i nanoszone w postaci zakodowanej na oddzielną kartę perforowaną. Po wprowadzeniu do pamięci operacyjnej pytań, a z kolei zapisów z taśmy magnetycznej, następowało porównywanie jednych i drugich danych. W przypadku zgodności deskryptorów danej publikacji z deskryptorami zadanymi w pytaniu następował zapis odpowiedzi na taśmie magnetycznej. Spis odpowiedzi na wszystkie pytania sporządzany był w formie wydruku. W przypadku niewystępowania w zbiorze publikacji, odpowiadającej zadanym deskryptorom, drukowana była odpowiedź: "na dany temat katalog nie zawiera ani jednej pozycji".

W toku eksperymentu stwierdzono, że na jedno pytanie przypada od 0 do 50 publikacji, a czas odpowiedzi wynosi średnio około 2 min.

Przykład 2.

Następny przykład dotyczy zautomatyzowanego systemu informacyjnego, wykorzystywanego do stawiania diagnoz lekarskich.

W wyniku długotrwałych prac badawczych zgromadzono dotychczas bardzo duże ilości informacji w dziedzinie leczenia chorych. Każda choroba charakteryzuje się zbiorem specyficznych dla niej symptomów, które występują rzadziej lub częściej, z mniejszym lub większym nasileniem, a więc dla danej choroby przyjmują różne "wagi".

Lekarz, badając chorego, nie jest w stanie, ze względu na ograniczony czas bądź brak doświadczenia, wykorzystać wszystkich wiadomości, jakimi dysponuje medycyna w dziedzinie ustalania diagnoz na podstawie znanej grupy symptomów i ich wag. Funkcję tę może spełnić szybciej i z większym prawdopodobieństwem elektroniczna maszyna cyfrowa. Zbiór informacji, przechowywany w maszynie, który można by nazwać zbiorem diagnostycznym, może być zbudowany na podstawie zero - jedynkowego zbioru prostokątnego, w którym choroby są elementami informacji, a symptomy o odpowiednich wagach - deskryptorami.

Prócz ustalania diagnoz lekarskich, elektroniczna maszyna cyfrowa może być wykorzystana do gromadzenia wszelkich informacji, składających się na historię choroby danego pacjenta, co w sumie pozwala jej sformułować optymalny plan leczenia chorego.

Wyszukiwanie informacji przy złagodzonych kryteriach

Rozpatrywany dotychczas przypadek wyszukiwania informacji wymagał spełnienia najbardziej ostrego kryterium, które polegało na znalezieniu wszystkich elementów informacji związanych z wszystkimi zadanymi deskryptorami. Wielokrotnie zadanie takie jest niewykonalne i wówczas stosuje się wyszukiwanie informacji przy kryteriach złagodzonych. Istnieje szereg przypadków tego rodzaju wyszukiwania, jak np. wyszukiwanie elementów związanych z k dowolnymi deskryptorami spośród n deskryptorów zadanych, przy czym k może się zmieniać od n do 1; wyszukiwanie elementów związanych jednocześnie z pewną wyróżnioną grupą spośród deskryptorów zadanych /najbardziej ważnych/ i w miarę możliwości z największą ilością pozostałych równorzędnych deskryptorów; wyszukiwanie elementów związanych tylko z jednym dowolnym lub większą liczbą deskryptorów wyróżnionych itp.

Przykładowo rozpatrzeniu zostanie pierwszy z przytoczonych przypadków wyszukiwania informacji.

Jeśli nie ma ani jednego elementu informacji, związanego z **WSZY-**stkiemi n zadanymi równorzędnymi deskryptorami, wówczas prowadzi się wyszukiwanie elementów informacji związanych z dowolnymi $m - 1/$ zadanymi deskryptorami. Jeśli i ten rodzaj wyszukiwania jest niewykonalny, wtedy poszukuje się elementów informacji związanych z dowolnymi $n - 2/$ zadanymi deskryptorami itd.

W tym przypadku najbardziej racjonalna jest metoda wyszukiwania informacji oparta na sortowaniu prostym, które może być realizowane za pomocą środków ręcznych, mechanicznych jak również automatycznych.

Sortowanie proste polega na tym, że wszystkie elementy informacji dzieli się na dwie grupy względem p-owego zadanego deskryptora D , tzn. w jednej grupie znajdują się elementy związane z deskryptorem D , w drugiej - elementy nie związane z nim.

Metoda ta rozpatrzona zostanie na przykładzie 3 zadanых deskryptorów: D_1, D_2, D_3 . Przebieg poszczególnych sortowań prostych zilustrowany jest na rys. 18.

Przy pierwszym sprawdzeniu przeprowadza się tyle sortowań prostych, ile jest zadaných deskryptorów. Przy każdym następnym sprawdzeniu liczba sortowań maleje o 1.

Liczba sprawdzeń równa jest ilości zadaných deskryptorów. Te elementy informacji, które związane są z D_1 , oznaczone zostały na rysunku przez E_{D_1} , te, które są związane z D_1 i D_2 - przez $E_{D_1 D_2}$ itd. Elementy nie związane z danym deskryptorem oznaczone są znakiem negacji

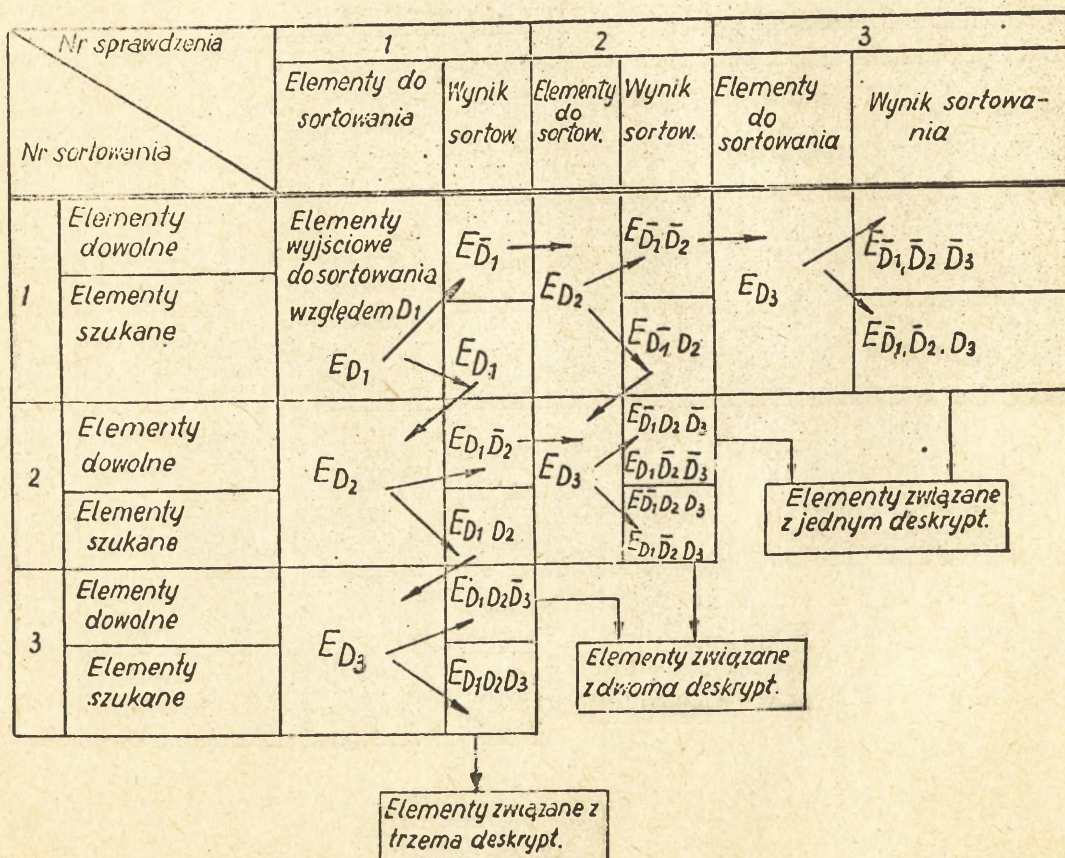
$/E_{D_1 D_2 D_3}$ oznacza elementy, które związane są jednocześnie z deskryptorami D_1 i D_3 , lecz nie są związane z $D_2/$.

Wyszukiwanie informacji przebiega następująco. Pierwsze sprawdzenie polega na ustaleniu, czy istnieją elementy informacji związane z wszystkimi trzema deskryptorami. W wyniku pierwszego sortowania elementów wyjściowych względem D_1 , w ramach pierwszego sprawdzenia, podlegają one podziałowi na 2 grupy. Jedna z nich posiada elementy informacji związane z D_1 i podlega dalszemu sortowaniu w tym samym sprawdzaniu, natomiast druga - nie związana z deskryptorem D_1 , wykorzystywana jest przy następnym sprawdzeniu.

Przy dalszych sortowaniach pierwszego sprawdzenia postępuje się analogicznie. Po drugim sortowaniu wydziela się grupę elementów zwią-

zanych jednocześnie z deskryptorami D_1 i D_2 oraz grupę elementów związanych z deskryptorem D_1 i nie związanych z deskryptorem D_2 . Po trzecim sortowaniu pierwszego sprawdzenia otrzymuje się dwie grupy elementów, z których jedna związana jest jednocześnie z deskryptorami D_1 , D_2 i D_3 , a druga związana jest jednocześnie z deskryptorami D_1 i D_2 , a nie jest związana z deskryptorem D_3 .

Jeżeli wynik pierwszego sprawdzenia jest negatywny, tzn. nie ma ani jednego elementu informacji, związanego jednocześnie z deskryptorami D_1, D_2, D_3 , wówczas łagodzi się warunki wyszukiwania i przystępuje do drugiego sprawdzenia. Drugie sprawdzenie pozwala stwierdzić, czy wśród elementów znajdują się takie, które związane są przynajmniej z dwoma dowolnymi deskryptorami. Jako pierwsze prowadzi się sortowanie elementów $E_{\bar{D}_1}$ /z pierwszego sprawdzenia/ względem deskryptora D_2 . Wyodrębnioną grupę elementów $E_{\bar{D}_1 D_2}$ łączy się z grupą elementów $E_{D_1 \bar{D}_2}$ /z drugiego sortowania pierwszego sprawdzenia/ i sortuje względem deskryptora D_3 . Dalej postępuje się analogicznie.



Rys. 18.

O ile wynik drugiego sprawdzenia jest również negatywny, wówczas prowadzi się trzecie sprawdzenie, przebiegające analogicznie, jak drugie. Sprawdzenie to powinno dać w wyniku pojedynczego sortowania wszystkie elementy informacji, związane przynajmniej z jednym deskryptorem.

ZAKOŃCZENIE

W zakończeniu należy jeszcze raz podkreślić, że omówione metody wyszukiwania informacji za pomocą różnych środków technicznych dotyczą przypadku, gdy wyszukiwanie prowadzone jest w zbiorach, w których informacja wyrażona jest za pomocą języka deskryptorowego, tzn. za pomocą elementów informacji i związanych z nimi deskryptorów.

Skrypt nie wyczerpuje zagadnień związanych z wyszukiwaniem informacji. Zwracając uwagę głównie na metody wyszukiwania informacji, potraktowano powierzchownie zagadnienie tworzenia języków deskryptorowych, jak również zasady optymalnego kodowania zbiorów informacji. Zagadnienia te są bowiem na tyle złożone, że wymagałyby odrębnego omówienia.

BIBLIOGRAFIA:

1. P. Ledli: Programmiowanie i ispolzowanie wyczyslitielnych maszin.
2. A.I. Michajłow: Osnovy naucznoi informacii.
3. I. Pijacidiewskij: Informacjonnyje sistiemy w tiechnieke i ekonomii-
kie.
4. Batrakov: Błektronnyje cifrowyje masziny dla reszenia informacjonno-
logiczeskich zadacz.

Wyk. w 50 egz.

Egz.nr 1- 50-B.Szk.-dział jawny

Wyk. Łęcki, ppłk

Druk JD, dnia 3.10.68r.

nr ks. 2245/2428/WW.

Druk ASG-0-XV-3895

