

Temat 6

Wojna morska – algorytmy przeszukiwania

Streszczenie

Wyszukiwanie informacji w wielkich zbiorach danych wymagają często użycia komputerów. Wymaga to ciągłego doskonalenia szybkich i efektywnych metod przeszukiwania. W czasie tych zajęć pokazane zostaną trzy różne metody: przeszukiwanie liniowe, przeszukiwanie binarne i haszowanie.

Wiek

- ✓ 9 i więcej

Materiały

Każde dziecko będzie potrzebować:

- ✓ Karty pracy
 - 1A, 1B for gry 1
 - 2A, 2B for gry 2
 - 3A, 3B for gry 3
- ✓ W razie potrzeby można użyć alternatywnych wersji 1A', 1B', 2A', 2B', 3A', 3B'.

Wojna morska

Wprowadzenie

1. Wybierz grupę dzieci, które następnie stają przed resztą klasy. Daj każdemu z dzieci (w przypadkowej kolejności) kartkę z zapisaną na niej liczbą. Liczby muszą być ukryte przed resztą uczniów.
2. Daj jednemu z pozostałych dzieci pojemnik, w którym znajduje się kilka cukierków. Jego zadaniem jest znaleźć daną liczbę. Może ono „płacić” za spojrzenie na konkretną kartę. Jeśli uda się odnaleźć poszukiwaną liczbę zanim wydane zostaną wszystkie cukierki, to resztę dziecko zabiera dla siebie.
3. Jeśli potrzeba, można wyżej opisaną część powtórzyć.
4. Następnie przetasuj kartki i rozdaj je ponownie. Dzieci powinny jednak tym razem ustawić się w kolejności „od najmniejszej do największej liczby”. W tym momencie uczniowie niech powtórzą proces przeszukiwania.

W przypadku gdy liczby są posortowane, rozsądną strategią jest „wykluczenie” połowy uczniów z dalszych przeszukiwań za pomocą tylko jednego „płacenia” za ujawnienie liczby dziecku, które jest w środku. Powtarzając tę strategię jeszcze co najwyżej dwukrotnie, dziecko poszukujące danej liczby znajdzie ją. Uczniowie z całą pewnością zauważą efektywność tej strategii.

Zadanie

Dzieci mogą wejść w rolę komputera przeszukującego zbiór danych, biorąc udział w grze „wojna morska”. W czasie gry powinni starać się przemyśleć stosowaną strategię.

Wojna morska – przeszukiwanie liniowe

Przeczytaj uczniom poniższą instrukcję

1. Podzielcie się na pary. Jedna z osób otrzyma kartę 1A, druga – 1B. Nie pokazujcie kart sobie nawzajem!
2. Każda z osób wybiera (zaznacza) jeden ze statków, znajdujących się w górnej części karty i przekazuje partnerowi informację o liczbie, którą statek jest oznaczony.
3. Następnie rozpoczyna się odgadywanie położenia statku przeciwnika. (Jedna osoba podaje literę-nazwę statku, a druga podaje liczbę, którą oznaczony jest ten statek.)
4. Wynikiem liczbowym uzyskanym przez daną osobę jest liczba prób („strzałów”), które były potrzebne do zlokalizowania statku drugiej osoby.

(Karty 1A' and 1B' rozdajemy tylko wtedy, gdy dzieci chciałyby rozegrać dodatkową grę lub wtedy, gdy „przypadkowo” dzieci zobaczyły kartę partnera. Karty 2A', 2B' i 3A', 3B' nie dotyczą tej gry.)

Dyskusja

1. Jakie uzyskaliście wyniki?
2. Jakim najmniejszym a jakim największym wynikiem może zakończyć się ta gra? (Są to odpowiednio 1 i 26 punktów, oczywiście przy założeniu, że dzieci nie wybierały żadnego ze statków dwukrotnie. Ta gra ilustruje dobrze metodę przeszukiwania zwaną „przeszukiwaniem liniowym”, ponieważ wiąże się ze sprawdzaniem wszystkich pozycji, jedna po drugiej.)

Wojna morska – przeszukiwanie binarne

Instrukcje

Instrukcje do tej wersji gry są takie same jak do poprzedniej. Należy jednak najpierw wyjaśnić dzieciom, że liczby, którymi opisane są statki są uporządkowane od najmniejszej do największej.

1. Podzielcie się na pary. Jedna z osób otrzyma kartę 1A, druga – 1B. Nie pokazujcie kart sobie nawzajem!
2. Każda z osób wybiera (zaznacza) jeden ze statków, znajdujących się w górnej części karty i przekazuje partnerowi informację o liczbie, którą statek jest oznaczony.
3. Następnie rozpoczyna się odgadywanie położenia statku przeciwnika. (Jedna osoba podaje literę-nazwę statku a druga podaje liczbę, którą oznaczony jest ten statek.)
4. Wynikiem liczbowym uzyskanym przez daną osobę jest liczba prób (strzałów), które były potrzebne do zlokalizowania statku partnera gry.

Dyskusja

1. Jakie uzyskaliście wyniki?
2. Jakiej strategii używaliście?
(Pytanie skierowane do tych, którzy zdobyli najniższy wynik liczbowy.)
3. Który ze statków powinien być wskazywany jako pierwszy podczas gry? (Ten statek, który znajduje się w środku – po jego wskazaniu wiemy już, w której linii znajduje się poszukiwany statek.) Który należy wskazać w kolejnym kroku? (Ponownie najlepszą strategią jest wskazywanie statku, który znajduje się w środku grupy.)
4. Jaka będzie liczba prób („strzałów”) w przypadku stosowania ww. strategii?
(Co najwyżej pięć.)

Przedstawiona wyżej metoda jest nazywana przeszukiwaniem binarnym (przeszukiwaniem przez połowienie), ponieważ stosuje się w niej podział zbioru na dwie części.

Wojna morska – przeszukiwanie z haszowaniem

Instrukcje

1. Każdy otrzymuje kartę i przekazuje partnerowi informację o liczbie, którą wybrany przez niego statek jest oznaczony.
2. W tej wersji gry można łatwo ustalić kolumnę (od 0 do 9), w której poszukiwany statek się znajduje. Po prostu należy ustalić sumę cyfr liczby, którą statek jest oznaczony – ostatnia cyfra tej sumy odpowiada wspomnianej kolumnie. Dla przykładu: aby zlokalizować statek opisany liczbą 2345, dodajemy cyfry tej liczby $2+3+4+5$ i otrzymujemy liczbę 14. Cyfra jedności to 4, więc statek musi znajdować się w kolumnie nr 4. Po ustaleniu kolumny pozostaje odgadnąć, który ze statków w tej kolumnie jest tym poszukiwanym.

(Zastosowana w oznaczaniu statków technika nazywana jest „haszowaniem”, ponieważ na podstawie kilku cyfr tworzona jest jedna cyfra-skrót. Angielski czasownik „hash” dosłownie oznacza „posiekać”.)

3. Rozpoczyna się gra, w czasie której należy zastosować ww. strategię. Można grę powtórzyć kilka razy, używając tej samej karty i wybierając statek w innej kolumnie.

(W przypadku tej wersji gry, inaczej niż w poprzednich, używamy kart 3A' i 3B' jako pary kart, ponieważ wzorce-skróty statków w kolumnach muszą odpowiadać sobie.)

Dyskusja

1. Zbierz i omów uzyskane przez uczniów wyniki.
2. Które statki można najszybciej znaleźć? (Te, które są jedynymi w kolumnie.) Które statki trudniej znaleźć? (Te, które znajdują się w bardziej licznych kolumnach.)
3. W przypadku której wersji gry proces przeszukiwania jest najszybszy? Dlaczego?

Jakie są zalety każdej z trzech przedstawionych metod przeszukiwania?

(Druga strategia jest szybsza od pierwszej, ale w przypadku pierwszej statki nie muszą być uporządkowane. Trzecia strategia jest zazwyczaj szybsza od dwóch pierwszych, chyba że wszystkie statki znajdują się w tej samej kolumnie – wtedy trzecia strategia jest tak samo powolna jak pierwsza.)

Zadania dodatkowe

1. Dzieci mogą przygotować własne karty dla każdej z wersji gier. W przypadku drugiej gry muszą pamiętać o tym, by liczby były w porządku od najmniejszej do największej. Można zapytać dzieci o to, jak stworzyć „najłatwiejszą” odmianę gry z haszowaniem. (Najłatwiejsza będzie taka z równą liczbą statków w każdej kolumnie. Najtrudniejsza – gdy wszystkie statki znajdują się w tej samej kolumnie.)
2. Co zdarzy się w przypadku gdy szukanego statku po prostu nie ma? (W wersji przeszukiwania liniowego uzasadnienie tego faktu wymaga 26 prób. W wersji przeszukiwania binarnego wystarczy pięć prób. W wersji z haszowaniem liczba prób zależy od liczby statków w odpowiedniej kolumnie.)
3. Jaka liczba prób (strzałów) w wersji przeszukiwania binarnego byłaby wystarczająca do rozstrzygnięcia gry w przypadku 100 statków (6-7), 1000 statków (9-10), miliona (19-20)?
(Uczniowie powinni zauważyć, że liczba prób-strzałów wzrasta bardzo powoli w porównaniu z liczbą statków: podwojenie liczby statków wymaga tylko jednej próby więcej niż wcześniej, tj. liczba prób rośnie proporcjonalnie do logarytmu liczby statków.)

9058	7169	3214	5891	4917	2767	4715	674	8088	1790	8949	13	3014
A	B	C	D	E	F	G	H	I	J	K	L	M
8311	7621	3542	9264	450	8562	4191	4932	9462	8423	5063	6221	2244
N	O	P	Q	R	S	T	U	V	W	X	Y	Z

A	B	C	D	E	F	G	H	I	J	K	L	M
N	O	P	Q	R	S	T	U	V	W	X	Y	Z

1A

1630	9263	4127	405	4429	7113	3176	4015	7976	88	3465	1571	8625
A	B	C	D	E	F	G	H	I	J	K	L	M
2587	7187	5258	8020	1919	141	4414	3056	9118	717	7021	3076	3336
N	O	P	Q	R	S	T	U	V	W	X	Y	Z

A	B	C	D	E	F	G	H	I	J	K	L	M
N	O	P	Q	R	S	T	U	V	W	X	Y	Z

1B

163	445	622	1410	1704	2169	2680	2713	2734	3972	4208	4871	5031
A	B	C	D	E	F	G	H	I	J	K	L	M
5283	5704	6025	6801	7440	7542	7956	8094	8672	9137	9224	9508	9663
N	O	P	Q	R	S	T	U	V	W	X	Y	Z

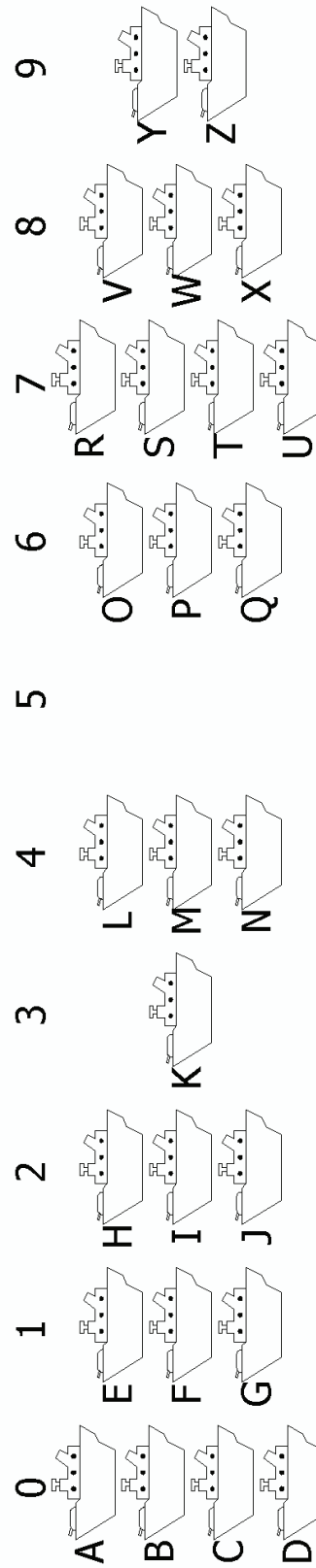
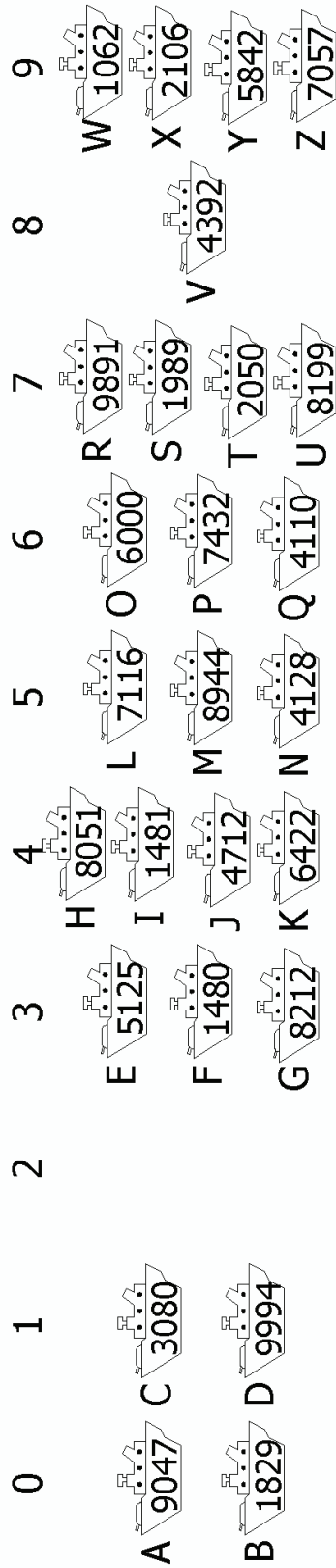
A	B	C	D	E	F	G	H	I	J	K	L	M
N	O	P	Q	R	S	T	U	V	W	X	Y	Z

2A

33	183	730	911	1927	1943	2200	2215	3451	3519	4055	5548	5655
A	B	C	D	E	F	G	H	I	J	K	L	M
5785	5897	5905	6118	6296	6625	6771	6831	7151	7806	8077	9024	9328
N	O	P	Q	R	S	T	U	V	W	X	Y	Z

A	B	C	D	E	F	G	H	I	J	K	L	M
N	O	P	Q	R	S	T	U	V	W	X	Y	Z

2B



3A

0	A 9308 B 1478 C 8417 D 9434	1	E 6519 F 2469 G 5105	2	H 1524 I 8112 J 2000	3	K 4135	4	L 9050 M 1265 N 5711	5		6	O 4200 P 7153 Q 6028	7	R 3121 S 9503 T 1114 U 7019	8	V 2385 W 5832 X 1917	9	Y 1990 Z 2502
---	--------------------------------------	---	----------------------------	---	----------------------------	---	--------	---	----------------------------	---	--	---	----------------------------	---	--------------------------------------	---	----------------------------	---	------------------

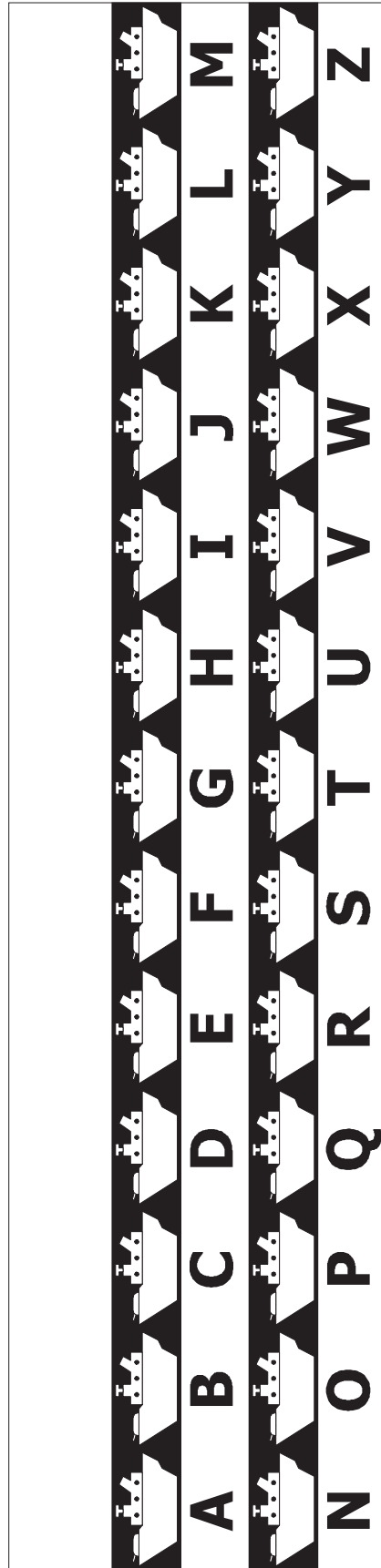
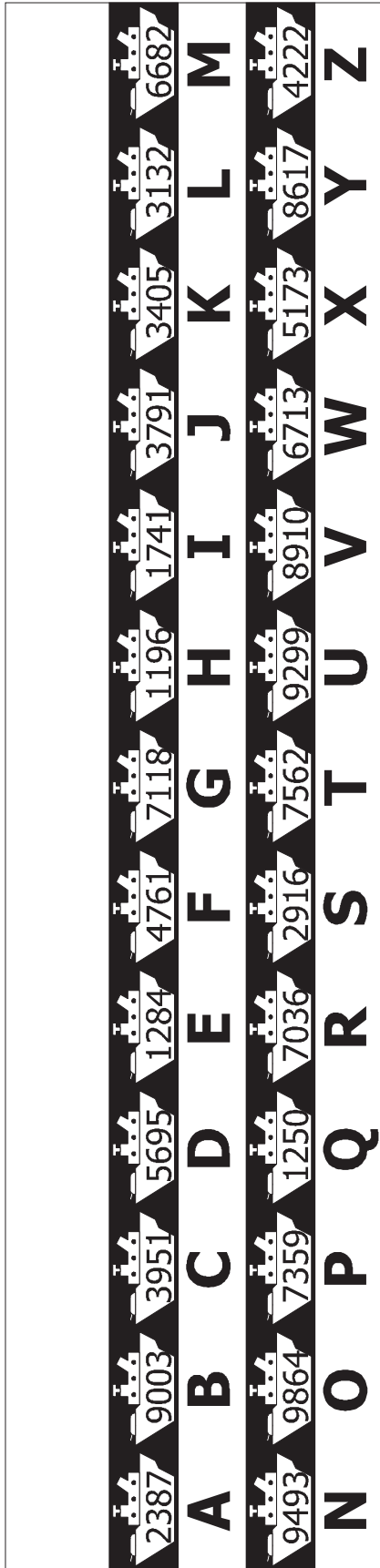
0	A B	1	C D	2		3	E F G	4	H I J K	5	L M N	6	O P Q	7	R S T U	8	V	9	W X Y Z
---	--------	---	--------	---	--	---	-------------	---	------------------	---	-------------	---	-------------	---	------------------	---	---	---	------------------

3B

6123	1519	9024	5164	2038	2142	7156	9974	9375	7104	1004	1023	5108
A	B	C	D	E	F	G	H	I	J	K	L	M
1884	3541	5251	4840	3289	3654	2480	5602	8965	4053	2405	2304	1959
N	O	P	Q	R	S	T	U	V	W	X	Y	Z

A	B	C	D	E	F	G	H	I	J	K	L	M
N	O	P	Q	R	S	T	U	V	W	X	Y	Z

1A'

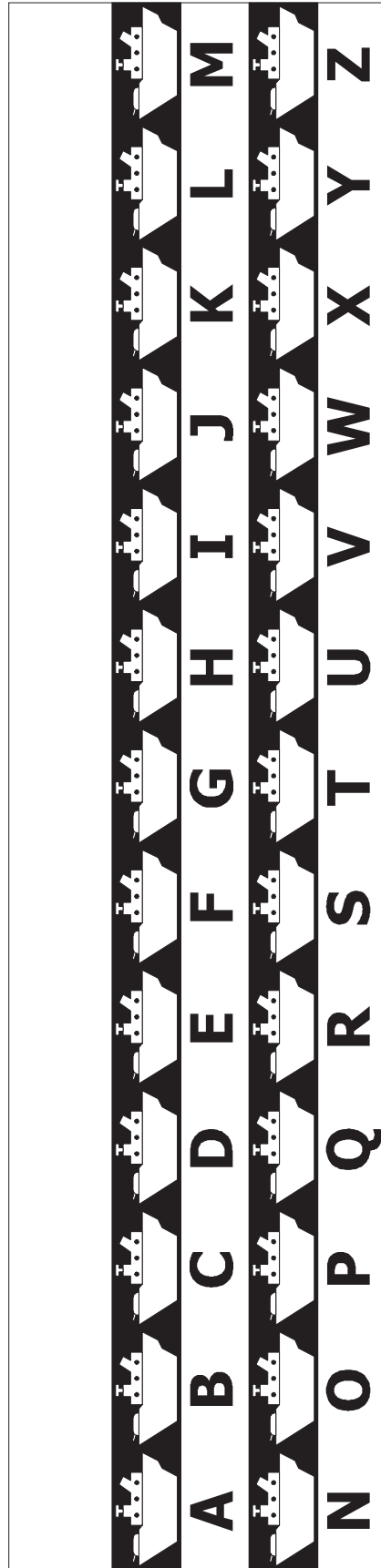
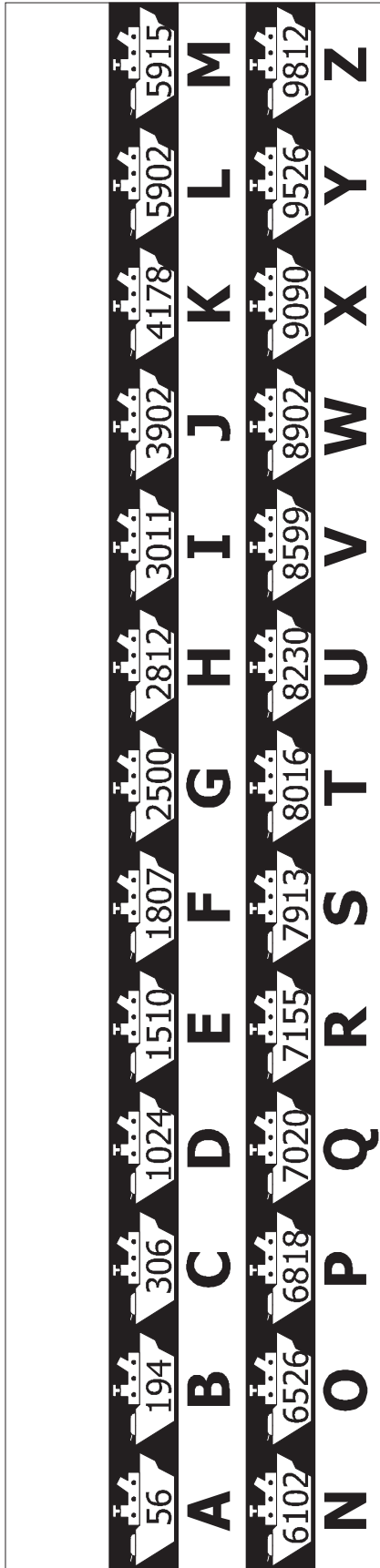


1B'

28	326	943	1321	1896	2346	2430	2929	3106	3417	4128	4717	4915
A	B	C	D	E	F	G	H	I	J	K	L	M
5123	5615	6100	7015	7120	7695	7812	8103	8719	9020	9608	9713	9911
N	O	P	Q	R	S	T	U	V	W	X	Y	Z

A	B	C	D	E	F	G	H	I	J	K	L	M
N	O	P	Q	R	S	T	U	V	W	X	Y	Z

2A'



2B'

0	A 1982	B 7841							
1	C 6113	D 1055							
2									
3	E 9121	F 1011	G 2984						
4	H 5009	I 2651	J 1751	K 4848					
5	L 1248	M 1716	N 2148						
6	O 2004	P 5173	Q 2806						
7	R 9369	S 1321	T 3004	U 7190					
8		V 3285							
9	W 9172	X 2052	Y 6012	Z 7525					

0	A	B	C	D					
1	E	F	G						
2	H	I	J						
3			K						
4	L	M	N						
5									
6	O	P	Q						
7	R	S	T	U					
8	V	W	X						
9	Y	Z							

3A'

0	A 8615 B 7003 C 1991 D 6211	1	E 1361 F 7644 G 5600	2	H 7726 I 9003 J 5557	3	K 3000	4	L 1814 M 2002 N 8844	5		6	O 9656 P 4002 Q 1221	7	R 6993 S 3121 T 4300 U 1907	8	V 8208 W 9423 X 4176	9	Y 2917 Z 4122
---	--------------------------------------	---	----------------------------	---	----------------------------	---	--------	---	----------------------------	---	--	---	----------------------------	---	--------------------------------------	---	----------------------------	---	------------------

0	A B	1	C D	2		3	E F G	4	H I J K	5	L M N	6	O P Q	7	R S T U	8	V	9	W X Y Z
---	--------	---	--------	---	--	---	-------------	---	------------------	---	-------------	---	-------------	---	------------------	---	---	---	------------------

3B'

O co w tym wszystkim chodzi?

Komputery przetwarzają duże ilości informacji i dlatego potrzebne są szybkie metody ich przeszukiwania. Z jednym z największych problemów tego rodzaju muszą zmierzyć się serwery wyszukiwarek internetowych, które w ciągu ułamka sekundy muszą dotrzeć do informacji zgromadzonej w jej archiwach. Dana, której wyszukiwaniem zajmuje się system komputerowy (np. słowo, liczba odczytana z kodu kreskowego, czy nazwisko autora) nazywa się *kluczem przeszukiwania*.

Komputery potrafią przetwarzać informacje z bardzo dużą szybkością, i można by pomyśleć, że w celu znalezienia konkretnej informacji powinny po prostu za każdym razem rozpoczynać przeszukiwanie od pierwszej danej zapisanej w pamięci, jak to zostało zapisane w wersji gry z przeszukiwaniem liniowym. Ta metoda jest jednak bardzo powolna – nawet dla komputerów. Dla przykładu: Wyobraźmy sobie supermarket, na którego półkach znaleźć można 10 000 różnych produktów. Kiedy przy kasie skanowany jest kod kreskowy, system komputerowy musi przeszukać nawet 10 000 liczb, zanim znajdzie nazwę produktu i jego cenę. Nawet jeśli sprawdzenie jednego kodu trwałoby tylko 0,001 sekundy, przeszukanie całej listy wymagałoby nawet 10 sekund. Jak długo trwałoby wówczas podliczenie ceny artykułów spożywczych dla jednej rodziny, która robi zakupy!

Lepszą strategią jest przeszukiwanie binarne (przez połowienie). W tym przypadku przeszukiwany jest zbiór uporządkowany. Sprawdzenie elementu środkowego na liście elementów pozwala już stwierdzić w której połowie zbioru znajduje się poszukiwany element (klucz wyszukiwania). Proces ten jest powtarzany aż do momentu znalezienia tego elementu. Wróćmy do przykładu supermarketu: 10 000 elementów listy produktów można przeszukać za pomocą co najwyżej 14 prób, co trwać będzie zaledwie 0,02 sekundy.

Trzecia strategia nosi nazwę przeszukiwania z haszowaniem. W tym przypadku na kluczu wyszukiwania wykonywana jest pomocnicza operacja, która ma pozwolić na ograniczenie zbioru przeszukiwania. Dla przykładu: Jeśli przeszukiwane mają być numery telefonów, można zastosować haszowanie oparte o tablicę indeksowaną przez reszty z dzielenia sumy wszystkich cyfr tworzących numer telefonu przez 11. Klucz haszowania (indeks tablicy), podobnie jak cyfry kontrolne omówione w czasie zajęć nr 4, to wartość, która jest zależna od właściwych danych, które są przetwarzane. Efekt jest taki, że zazwyczaj komputer odnajdzie poszukiwany element niemal natychmiast. Jest małe prawdopodobieństwo, że większość numerów telefonów „trafi” do tej samej „kolumny” tabeli haszującej i komputer będzie musiał przeszukiwać duży zbiór.

Programiści, w przypadku zbiorów danych, na których wykonywane będą często operacje przeszukiwania, zazwyczaj używają jednej z odmian strategii haszowania. Chyba, że z jakiegoś powodu jest ważne przechowywać w pamięci dane w postaci uporządkowanej albo trudne do zaakceptowania jest nawet sporadyczne długie oczekiwanie na odpowiedź programu wyszukującego.