

26.04.2012 Gliwice

Laboratorium

Podstawy Informatyki

Metody dostępu do danych

Autor:	Jonatan Dragon,
	Informatyka, sem. 2, gr. 6
Prowadzący:	dr inż. Ewa Płuciennik-Psota

Zad. 1 Przygotować zbiory z danymi testowymi.

A100.dat:	31 : 668	64 : 233	97 : 838	2 : 378	11 : 603
0 : 173	32 : 910	65 : 609	98 : 819	3 : 530	12 : 518
1 : 701	33 : 776	66 : 689	99 : 548	4 : 684	13 : 276
2 : 364	34 : 59	67 : 142		5 : 671	14 : 889
3 : 933	35 : 837	68 : 304	B20.dat:	6 : 152	15 : 816
4 : 980	36 : 823	69 : 962	0 : 946	7 : 895	16 : 700
5 : 155	37 : 868	70 : 555	1 : 726	8 : 612	17 : 311
6 : 582	38 : 147	71 : 559	2 : 522	9 : 930	18 : 409
7 : 100	39 : 366	72 : 726	3 : 233	10 : 154	19 : 680
8 : 864	40 : 696	73 : 285	4 : 643	11 : 461	20 : 70
9 : 83	41 : 288	74 : 859	5 : 149	12 : 227	21 : 177
10 : 160	42 : 32	75 : 690	6 : 866	13 : 23	22 : 542
11 : 568	43 : 273	76 : 250	7 : 643	14 : 61	23 : 987
12 : 643	44 : 299	77 : 214	8 : 299	15 : 27	24 : 982
13 : 905	45 : 993	78 : 956	9 : 848	16 : 376	25 : 301
14 : 549	46 : 778	79 : 931	10 : 284	17 : 116	26 : 465
15 : 142	47 : 950	80 : 663	11 : 910	18 : 163	27 : 585
16 : 848	48 : 747	81 : 516	12 : 946	19 : 126	28 : 155
17 : 534	49 : 101	82 : 565	13 : 864		29 : 500
18 : 382	50 : 436	83 : 19	14 : 233	D31.dat:	30 : 139
19 : 78	51 : 779	84 : 149	15 : 864	0 : 848	
20 : 935	52 : 284	85 : 313	16 : 696	1 : 612	E5.dat:
21 : 897	53 : 254	86 : 62	17 : 696	2 : 457	0 : 501
22 : 260	54 : 90	87 : 458	18 : 848	3 : 411	1 : 518
23 : 925	55 : 266	88 : 81	19 : 299	4 : 50	2 : 50
24 : 193	56 : 634	89 : 492		5 : 928	3 : 982
25 : 896	57 : 810	90 : 741	C20.dat:	6 : 728	4 : 501
26 : 736	58 : 980	91 : 749	0 : 599	7 : 27	
27 : 429	59 : 604	92 : 866	1 : 139	8 : 501	
28 : 946	60 : 522	93 : 361		9 : 426	
29 : 859	61 : 894	94 : 269		10 : 367	
30 : 762	62 : 946	95 : 601			
	63 : 855	96 : 694			

Zad. 2 Wyszukiwanie sekwencyjne

Wyszukiwanie sekwencyjne	B20.dat	C20.dat
Dane nie posortowane	Min = 9, Max = 93, średnio 41,8	Min = 100, Max = 100, średnio 100
Dane posortowane	Min = 14, Max = 93, średnio 57,5	Min = 100, Max = 100, średnio 100

Powyżej przedstawiłem wyniki otrzymane przy wyszukiwaniu sekwencyjnym zbiorów danych B20 i C20 w zbiorze A100. W przypadku danych B20 posortowanie elementów spowodowało zwiększenie średniej ilości zapytań, co oznacza, że po posortowaniu dane B20 są bliżej maksymalnych wartości A100. Ze względu na brak danych C20 w zbiorze A100 ilość zapytań wyniosła ilości przeszukiwanych danych.

Zad. 3 Wyszukiwanie metodą podziałów dychotomicznych

B20.dat	C20.dat
Min = 3, Max = 7, średnio 6,1	Min = 6, Max = 7, średnio 6,85

W stosunku do metody sekwencyjnej w przypadku podziałów dychotomicznych można zaobserwować znaczny spadek liczby zapytań. Wadą tej metody jest konieczność posortowania danych przed wyszukiwaniem.

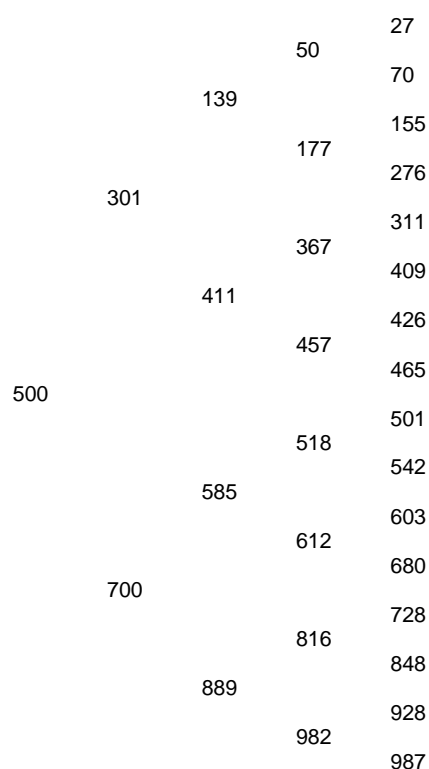
Zad. 4 Eksperymenty z drzewami binarnymi

Wyszukiwanie w drzewach binarnych	E5.dat
Bez wyważenia	Min = 4, Max = 6, średnio 4,6
Dokładnie wyważone	Min = 4, Max = 5, średnio 4,4

Średnia liczba wyszukań jak i ich maksymalna liczba, dla drzew wyważonych, jest nieznacznie mniejsza w porównaniu do drzew dokładnie wyważonych.

Zad. 5 Dokładnie wyważone drzewo binarne

1: 500	8: 50	15: 982	22: 426	29: 848
2: 301	9: 177	16: 27	23: 465	30: 928
3: 700	10: 367	17: 70	24: 501	31: 987
4: 139	11: 457	18: 155	25: 542	
5: 411	12: 518	19: 276	26: 603	
6: 585	13: 612	20: 311	27: 680	
7: 889	14: 816	21: 409	28: 728	



Zad. 6 Wyszukiwanie w B - drzewach

Rozmiar strony	B20.dat	C20.dat
4	Dostęp do pamięci: Min = 3, Max = 8, średnio 5,45 Dostęp do dysku: Min = 2, Max = 4, średnio 3,55	Dostęp do pamięci: Min = 5, Max = 8, średnio 6,75 Dostęp do dysku: Min = 4, Max = 4, średnio 4
8	Dostęp do pamięci: Min = 4, Max = 7, średnio 5,85 Dostęp do dysku: Min = 2, Max = 3, średnio 2,8	Dostęp do pamięci: Min = 5, Max = 8, średnio 6,6 Dostęp do dysku: Min = 3, Max = 3, średnio 3
16	Dostęp do pamięci: Min = 2, Max = 7, średnio 5,2 Dostęp do dysku: Min = 1, Max = 2, średnio 1,8	Dostęp do pamięci: Min = 6, Max = 8, średnio 6,8 Dostęp do dysku: Min = 2, Max = 2, średnio 2
32	Dostęp do pamięci: Min = 1, Max = 7, średnio 5,85 Dostęp do dysku: Min = 1, Max = 2, średnio 1,9	Dostęp do pamięci: Min = 6, Max = 7, średnio 6,7 Dostęp do dysku: Min = 2, Max = 2, średnio 2
64	Dostęp do pamięci: Min = 2, Max = 7, średnio 4,75 Dostęp do dysku: Min = 2, Max = 2, średnio 2	Dostęp do pamięci: Min = 6, Max = 7, średnio 6,75. Dostęp do dysku: Min = 2, Max = 2, średnio 2
128	Dostęp do pamięci: Min = 3, Max = 7, średnio 5,85 Dostęp do dysku: Min = 1, Max = 1, średnio 1	Dostęp do pamięci: Min = 6, Max = 7, średnio 6,8 Dostęp do dysku: Min = 1, Max = 1, średnio 1

Dla wszystkich rozmiarów strony, średnie wartości ilości dostępu do pamięci są podobne. Można więc wywnioskować, że rozmiar strony nie wpływa znacząco na tę wartość. Wyszukiwanie danych B20, które zawierają się w zbiorze A100, jak i wyszukiwanie danych C20 wydaje się najoptymalniejsze pod względem wykorzystania dysku dla rozmiaru strony równego 128. W tym wypadku tylko raz odczytujemy dane z dysku. Na podstawie tych informacji możemy stwierdzić, że wyszukiwanie w B-drzewach jest o znacznie wydajniejsze niż wyszukiwanie sekwencyjne.

Zad. 7 Wyszukiwanie w B* - drzewach

Rozmiar strony	B20.dat	C20.dat
4	Dostęp do pamięci: Min = 6, Max = 9, średnio 7,15. Dostęp do dysku: Min = 4, Max = 4, średnio 4	Dostęp do pamięci: Min = 5, Max = 8, średnio 6,5 Dostęp do dysku: Min = 4, Max = 4, średnio 4
8	Dostęp do pamięci: Min = 5, Max = 7, średnio 6,55 Dostęp do dysku: Min = 3, Max = 3, średnio 3	Dostęp do pamięci: Min = 5, Max = 8, średnio 6,8 Dostęp do dysku: Min = 3, Max = 3, średnio 3
16	Dostęp do pamięci: Min = 4, Max = 7, średnio 5,75 Dostęp do dysku: Min = 2, Max = 2, średnio 2	Dostęp do pamięci: Min = 6, Max = 7, średnio 6,4 Dostęp do dysku: Min = 2, Max = 2, średnio 2
32	Dostęp do pamięci: Min = 4, Max = 7, średnio 5,4 Dostęp do dysku: Min = 2, Max = 2, średnio 2	Dostęp do pamięci: Min = 6, Max = 7, średnio 6,5 Dostęp do dysku: Min = 2, Max = 2, średnio 2
64	Dostęp do pamięci: Min = 2, Max = 7, średnio 4,75 Dostęp do dysku: Min = 2, Max = 2, średnio 2	Dostęp do pamięci: Min = 6, Max = 7, średnio 6,65 Dostęp do dysku: Min = 2, Max = 2, średnio 2
128	Dostęp do pamięci: Min = 3, Max = 7, średnio 5,85 Dostęp do dysku: Min = 1, Max = 1, średnio 1	Dostęp do pamięci: Min = 6, Max = 7, średnio 6,8 Dostęp do dysku: Min = 1, Max = 1, średnio 1

Dla danych B20 wraz ze wzrostem rozmiaru strony zmniejsza się zarówno liczba dostępów do dysku, jak i do pamięci. W przypadku danych C20 teza ta jest prawdziwa tylko dla ilości dostępów do dysku. Najlepszym rozwiązaniem ponownie wydaje się być używanie dużego rozmiaru strony.

Zad. 8 Funkcja mieszająca

Wycięcie 3 cyfr klucza i normalizacja			
Funkcja rozwiązywania kolizji	Minimalny rozmiar	Rozmiar	Ilość kolizji
Sondowanie liniowe z krokiem 1	100	100	Min = 1, Max = 100, średnio 48,65
		150	Min = 1, Max = 100, średnio 48,65
		250	Min = 1, Max = 100, średnio 48,65
		300	Min = 1, Max = 100, średnio 48,65
Sondowanie liniowe z krokiem 7	100	100	Min = 1, Max = 100, średnio 48,65
		150	Min = 1, Max = 100, średnio 48,65
		250	Min = 1, Max = 100, średnio 48,65
		300	Min = 1, Max = 100, średnio 48,65
Podwójne mieszanie zależne	100	100	Min = 1, Max = 100, średnio 48,65
		150	Min = 1, Max = 100, średnio 48,65
		250	Min = 1, Max = 100, średnio 48,65
		300	Min = 1, Max = 100, średnio 48,65
Podwójne mieszanie niezależne	250	250	Min = 1, Max = 171, średnio 66,37
		400	Min = 1, Max = 123, średnio 53,76
		500	Min = 1, Max = 125, średnio 54,41
		700	Min = 1, Max = 128, średnio 53,75

Na podstawie uzyskanych danych możemy zauważyć, że dla pierwszych trzech funkcji rozwiązywania kolizji, ilość kolizji jest taka sama, i średnio wynosi 48,65. Oznacza to, że w tym wypadku, sposób rozwiązywania kolizji nie wpływa na ich ilość. Inaczej jest w funkcji podwójnego mieszania niezależnego. Ilość kolizji, jak i rozmiar tablicy wzrasta.

Podział, składanie i dzielenie			
Funkcja rozwiązywania kolizji	Minimalny rozmiar	Rozmiar	Ilość kolizji
Sondowanie liniowe z krokiem 1	100	100	Min = 1, Max = 83, średnio 8,05
		115	Min = 1, Max = 27, średnio 3,34
		150	Min = 1, Max = 10, średnio 1,55
		250	Min = 1, Max = 3, średnio 1,16
Sondowanie liniowe z krokiem 7	100	100	Min = 1, Max = 98, średnio 7,72
		106	Min = 1, Max = 43, średnio 4,22
		115	Min = 1, Max = 14, średnio 2,39
		150	Min = 1, Max = 9, średnio 1,69
Podwójne mieszanie zależne	151	151	Min = 1, Max = 5, średnio 1,68
		200	Min = 1, Max = 5, średnio 1,48
		300	Min = 1, Max = 7, średnio 1,36
		400	Min = 1, Max = 4, średnio 1,12
Podwójne mieszanie niezależne	106	106	Min = 1, Max = 16, średnio 2,39
		126	Min = 1, Max = 10, średnio 1,86
		156	Min = 1, Max = 8, średnio 1,46
		250	Min = 1, Max = 4, średnio 1,2

Funkcja mieszająca „Podział, składanie i dzielenie” znacznie zmniejsza ilość kolizji w porównaniu do poprzedniej metody. Dla pierwszych dwóch funkcji ilość kolizji gwałtownie maleje przy zwiększaniu rozmiaru tablicy.

Dzielenie przez rozmiar tablicy			
Funkcja rozwiązywania kolizji	Minimalny rozmiar	Rozmiar	Ilość kolizji
Sondowanie liniowe z krokiem 1	100	100	Min = 1, Max = 83, średnio 8,05
		105	Min = 1, Max = 32, średnio 2,67
		115	Min = 1, Max = 27, średnio 3,34
		150	Min = 1, Max = 10, średnio 1,55
Sondowanie liniowe z krokiem 7	100	100	Min = 1, Max = 98, średnio 7,72
		106	Min = 1, Max = 43, średnio 4,22
		115	Min = 1, Max = 14, średnio 2,39
		150	Min = 1, Max = 9, średnio 1,69
Podwójne mieszanie zależne	151	151	Min = 1, Max = 5, średnio 1,68
		200	Min = 1, Max = 5, średnio 1,48
		250	Min = 1, Max = 5, średnio 1,4
		300	Min = 1, Max = 7, średnio 1,36
Podwójne mieszanie niezależne	106	106	Min = 1, Max = 16, średnio 2,39
		150	Min = 1, Max = 10, średnio 1,61
		200	Min = 1, Max = 5, średnio 1,28
		300	Min = 1, Max = 3, średnio 1,17

Funkcja mieszająca „Dzielenie przez rozmiar tablicy” nie różni się znacząco od poprzedniej metody. Dla pierwszej, drugiej i czwartej funkcji rozwiązywania kolizji zwiększenie rozmiaru tablicy powoduje gwałtowny spadek ilości kolizji.

Mieszanie Fibonacciego			
Funkcja rozwiązywania kolizji	Minimalny rozmiar	Rozmiar	Ilość kolizji
Sondowanie liniowe z krokiem 1	100	100	Min = 1, Max = 98, średnio 6,94
		105	Min = 1, Max = 72, średnio 5,18
		115	Min = 1, Max = 52, średnio 3,04
		150	Min = 1, Max = 9, średnio 1,49
Sondowanie liniowe z krokiem 7	100	100	Min = 1, Max = 98, średnio 5,73
		106	Min = 1, Max = 33, średnio 3,43
		115	Min = 1, Max = 26, średnio 2,25
		150	Min = 1, Max = 7, średnio 1,51
Podwójne mieszanie zależne	151	151	Min = 1, Max = 9, średnio 1,83
		200	Min = 1, Max = 8, średnio 1,64
		300	Min = 1, Max = 7, średnio 1,29
		400	Min = 1, Max = 4, średnio 1,23
Podwójne mieszanie niezależne	100	106	Min = 1, Max = 18, średnio 2,69
		150	Min = 1, Max = 4, średnio 1,39
		200	Min = 1, Max = 4, średnio 1,28
		300	Min = 1, Max = 3, średnio 1,07

Funkcja mieszania Fibonacciego dla moich danych wydaje się być najlepsza. Występuje najmniej kolizji, a rozmiary tablicy pozostają prawie bez zmian.