Písomná skúška z ADŠ dňa 19. 12. 2012 – predtermín

1 (True or False) and Justify [20 bodov]

O každom tvrdení uveďte, či je pravdivé, a tromi vetami svoj názor zdôvodnite. Úplne správna odpoveď: 2.5 boda. Správna odpoveď bez zdôvodnenia: 0.5 boda. Nesprávna odpoveď: 0 bodov. Správna odpoveď s úplne zlým zdôvodnením: 0 bodov.

1. Ak v 99-prvkovom poli $470\times$ vymením dva rôzne, náhodne vybrané prvky, môžem na konci dostať ľubovoľnú permutáciu.
2. Z haldy sa dá vyrobiť binárny vyhľadávací strom (nie nutne vyvážený) v čase $O(n)$.
3. Ak má n-vrcholový binárny strom aspoň $n/2$ listov, tak je jeho hĺbka $O(\log n)$.
4. Každý binárny strom hĺbky $\leq 2\log_2 n$ je nutne vyvážený (v zmysle def. pre AVL stromy).
5. V halde s minimom v koreni pre ľubovoľné prvky x, y platí: ak $x > y$, tak x je aspoň tak hlboko ako y.
6. Pole obsahujúce n celých čísel z množiny $\{0, 1, \ldots, 47n - 1\}$ vieme usporiadať v čase asymptoticky lepšom ako $n \log n$.
7. Pole obsahujúce n celých čísel z množiny $\{0, 1, \ldots, n^{47} - 1\}$ vieme usporiadať v čase asymptoticky lepšom ako $n \log n$.
8. Keď do poľa veľkosti 1 000 000 zahešujem 10 000 prvkov, pravdepodobnosť aspoň jednej kolízie je väčšia ako $1/2$.

2 Pohľad zvnútra: stromy a haldy [5+5+5 bodov + 10 bonus]

a) Nakreslite *všetky* možné binárne stromy, pre ktoré platí: množina vrcholov je $\{1, 2, 3, 4, 5\}$, je to binárny vyhľadávací strom a zároveň aj minimová halda (t.j., každý vrchol spĺňa podmienku haldy). Zdôvodnite, prečo iné nevyhovujú.

b) Binárnu haldu s n prvkami vieme uložiť do poľa $A[0..n-1]$ tak, že pre vrchol uložený na indexe i sú jeho synovia (ak existujú) uložení na indexoch $2i + 1$ a $2i + 2$. Aké veľké pole (presne!) by sme potrebovali, ak by sme doň chceli vedieť rovnakým spôsobom uložiť ľubovoľný n-vrcholový binárny vyhľadávací strom?

c) Aké veľké pole (asymptoticky) by sme potrebovali v časti b), ak by bolo zaručené, že ten strom má hĺbku $\leq 2\log_2 n$?

d) BONUS: Aké veľké pole by sme potrebovali v časti b), ak by bolo zaručené, že ten strom je vyvážený?

3 Pohľad zvnútra: Niekoľko najväčších [6+14 bodov]

Neusporiadané pole $A[0..n-1]$ obsahuje prvky, ktoré vieme porovnávať. Chceme nájsť jeho $k = \lceil \sqrt{n} \rceil$ najväčších prvkov.
a) Dokážte, že každý program riešiaci túto úlohu musí mať časovú zložitosť $\Omega(n)$, t.j., lineárnu alebo horšiu.
b) Napíšte (naozaj do detailov) čo najefektívnejší program riešiaci túto úlohu. Odhadnite jeho časovú zložitosť.

Hint: Existuje veľa optimálnych riešení. Jedno vhodne použije haldu. Druhé si pole naseká na vhodne dlhé úseky. Tretie sa použitiu celého triedenia vyhne ešte iným spôsobom.

4 Pohľad zvonka: Porovnaj tri naraz [8+12 bodov]

V poli $A[0..n-1]$, pričom $n \geq 3$, máme navzájom rôzne prvky neznámeho typu. Prvky majú navzájom rôzne veľkosti, tie ale nepoznáme. Chceli by sme pole usporiadať podľa veľkosti prvkov. Jediný spôsob, ako vieme prvky porovnávať, je ale zvláštny: Máme funkciu, ktorej dáme tri *rôzne* indexy do poľa a dozvieme sa (v konštantnom čase), ktorý z tých troch prvkov je najmenší, ktorý je prostredný a ktorý najväčší.
a) Napíšte program, ktorý túto úlohu vyrieši s optimálnou časovou zložitosťou.
b) Dokážte, že asymptotická časová zložitosť vášho programu je optimálna.
(V oboch podúlohách je OK sa bez rozpisovania odvolávať na algoritmy a dôkazy z prednášok.)

5 Pohľad zvonka: DNA [5 × 5 bodov + 5 bonus]

Dve vlákna DNA na seba pasujú, ak majú rovnakú dĺžku a obsahujú komplementárne dusíkaté bázy (C ku G, A ku T). Teda napr. $GATTACA$ a $CTAATGT$ na seba pasujú. Toto môžeme zovšeobecniť. *Nekompatibilita* vláken α a β bude počet elementárnych zmien potrebných na to, aby α a β na seba pasovali. Povolené elementárne zmeny budeme uvažovať nasledovné: i) vynechanie niektorej bázy z vlákna, ii) vloženie novej bázy kamkoľvek do vlákna, iii) zmena ľubovoľnej bázy na inú.
Napr. nekompatibilita vláken $\alpha = GGGAGGT$ a $\beta = CTCCCCT$ je 3. Jeden optimálny spôsob úprav: z α vynecháme 4. bázu (A), potom do α na druhú pozíciu vložíme novú bázu A, a nakoniec v β poslednú bázu zmeníme z T na A. Dostaneme $\alpha' = GAGGGGT$ a $\beta' = CTCCCCA$ a tie už na seba pasujú.
V tejto úlohe sa skúsite dopracovať k efektívnemu algoritmu, ktorý pre dané dva reťazce spočíta ich nekompatibilitu.

a) Dokážte, že sa nekompatibilita žiadnej dvojice reťazcov nezmení, ak zakážeme elementárne operácie typu ii) – dovolíme teda len mazať bázy a meniť existujúce bázy na iné.

b) Napíšte (ako pseudokód alebo kus programu) rekurzívny algoritmus skúšajúci všetky možnosti ako postupne upravovať vlákna pomocou operácií typu i) a iii). Vhodné je začať napr. od konca: Ako môže vyzerať optimálna postupnosť úprav ak na seba posledné znaky pasujú? A ako, keď nepasujú?

c) Pridajte do predchádzajúceho algoritmu memoizáciu tak, aby vznikol algoritmus s časovou zložitosťou polynomiálnou od počtu vecí n. Odhadnite jeho časovú zložitosť.

d) Uveďte ekvivalentný algoritmus, ktorý túto úlohu rieši pomocou dynamického programovania.

e) Dokážte alebo vyvráťte správnosť pažravého algoritmu: Ak je niektorý reťazec prázdny, zmažem všetky písmená druhého. Ak na seba prvé písmená oboch pasujú, oba zmažem, ale nepočítam to medzi operácie. Ak nie a reťazce sú rôznej dĺžky, zmažem prvé písmeno dlhšieho. A ak na seba nepasujú prvé písmená dvoch rovnako dlhých reťazcov, zmením jedno z nich tak, aby už pasovali.

f) BONUS: V praxi čakáme, že budeme spracúvať hlavne dvojice vláken, ktoré na seba skoro pasujú. Ktorý z algoritmov v častiach c), d) a e) je najvhodnejší a prečo?