Osmé cvičení ADS 1

Příklad 1 (spodní odlad složitosti sloučení stromů): Dokažte, že pro množiny reprezentované binárními vyhledávacími stromy není možné sloučit v lepším než lineárním čase, a to ani v případě, kdy na vstupu dostanete dokonale vyvážené stromy a výstup nemusí být vyvážený.

Příklad 2 (výpočet frekvence): Je dán text rozdělený na slova. Chceme vypsat frekvenční slovník, tedy tabulku všech slov setříděných podle počtu výskytů.

Příklad 3 (podslova): Navrhněte algoritmus, který pro daný řetezec a hodnotu m najde všechny opakované výskyty slov délky m.

Příklad 4 (součet): Mějme množinu přirozených čísel a číslo x. Chceme zjistit, zda množina obsahuje dvojici prvků se součtem x.

Příklad 5 ((a,b)-stromy): Nevýhodou (a,b)-stromů je, že plýtvají pamětí – může se stát, že vrcholy jsou zaplněné jen z poloviny. Navrhněte úpravu, která zaručí zaplnění z alespoň $2/3$.

Příklad 5 (Bloomův filtr): *Bloomův filtr* je datová struktura pro přibližnou reprezentaci množiny. Skládá se z pole bitů $B[1\dots m]$ a hešovací funkce h, která prvkům univerza přiřazuje indexy v poli.

- Insert(x) nastaví $B[h(x)] = 1$.
- Member(x) otestuje, zda $B[h(x)] = 1$.

Vložme nyní do filtru nějakou n-prvkovou množinu M. Pokud $x \in M$, Member(x) vždy odpoví správně. Pokud se ale zeptáme na $x \notin M$, může se stát, že $h(x) = h(y)$ pro nějaké $y \in M$, a Member(x) odpoví špatně. Spočítejte, s jakou pravděpodobností se to pro dané m a n stane.

Využijte toho, že $1 + \alpha \leq e^{\alpha}$ pro každé $\alpha \in \mathbb{R}$.

Příklad 6 (optimalizace Bloomova filtru): Spolehlivost Bloomova filtru můžeme zvýšit tak, že si pořídíme k filtrů s různými hešovacími funkcemi. Insertb ude vkládat do všech, Member se zeptá všech a odpoví ano pouze tehdy, když se na tom všechny filtry shodnou. Je-li pravděpodobnost chyby jednoho filtru p, pak kombinace k filtrů chybuje s pravděpodobností pouhých p^k. Vymyslete, jak nastavit m a k pro případ, kdy chceme ukládat 10^6 prvků s pravděpodobností chyby nejvýše 10^{-9}. Minimalizujte spotřebu paměti.

Domácí úkol 10 (básníci, 10 bodů): Navrhněte datovou strukturu pro básníky, která si bude pamatovat slovník a bude umět hledat rýmy. Tedy pro libovolné zadané slovo najde jiné slovo ve slovníku, kterémá se zadaným co nejdelší společný suffix.

Domácí úkol 11 (medián, 15 bodů): Na vstupu postupně přicházejí čísla. Kdykoliv přijde další, vypište medián z posledních k čísel (to je číslo, které by bylo prostřední, kdybychom k-tici setřídili). Dosáhněte časové složitosti $O(\log k)$ na operaci.