

## Deváté cvičení ADS 1

**Příklad 1 (přesmyčky):** Jak reprezentovat slovník (množinu slov), abyste uměli rychle vyhledávat všechny přesmyčky zadaného slova?

**Příklad 2 (( $a, b$ )-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 3 (RB-stromy):** Odvoďte operace pro vkládání a mazání v RB-stromech. Je dobré vyjít z implementace (2, 3) nebo (2, 4)-stromů.

**Příklad 4 (Mimozemský narozeninový paradox.):** Mějme planetu s  $m$  dny v roce a skupinu  $n$  obyvatel této planety, kteří mají rovnoměrně náhodně narozeniny v tamním roce. Zkuste co nejlépe spočítat pravděpodobnost, že dva mimozemšťané v této skupině mají narozeniny ve stejný den.

**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.

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

Vložme nyní do filtru nějakou  $n$ -prvkovou množinu  $M$ . Pokud  $x \in M$ ,  $\text{Member}(x)$  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  $\text{Member}(x)$  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.  $\text{Insert}$  ude vkládat do všech,  $\text{Member}$  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 12 (join, 15 bodů):** Nechť  $X, Y$  jsou dva ( $a, b$ )-stromy takové, že všechny klíče v  $X$  jsou menší než všechny klíče v  $Y$ . Vymyslete algoritmus, který oba stromy v logaritmickeém čase sloučí do jednoho.